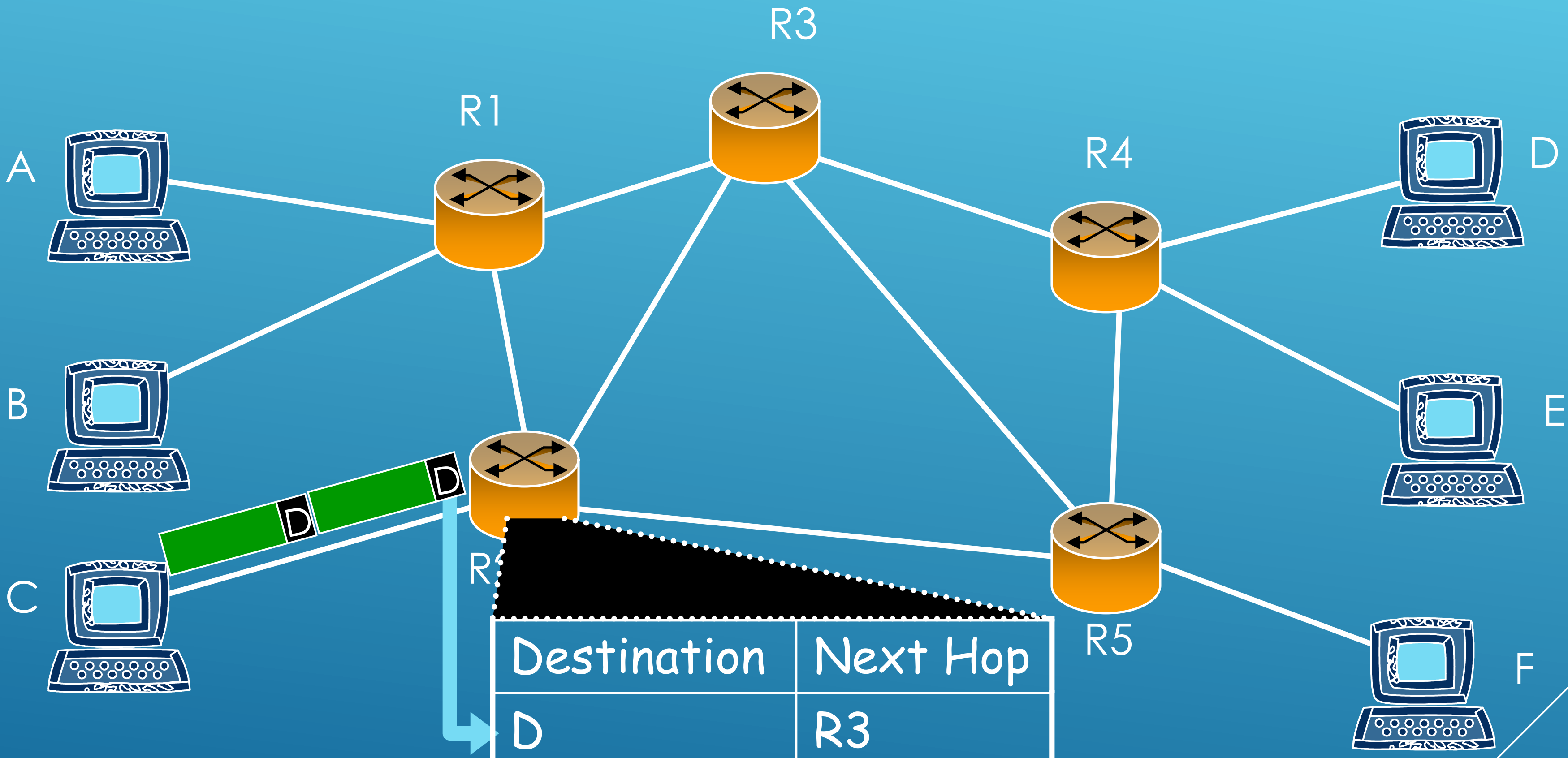
ROUTERS
IP TO MPLS TO CESR

OUTLINE

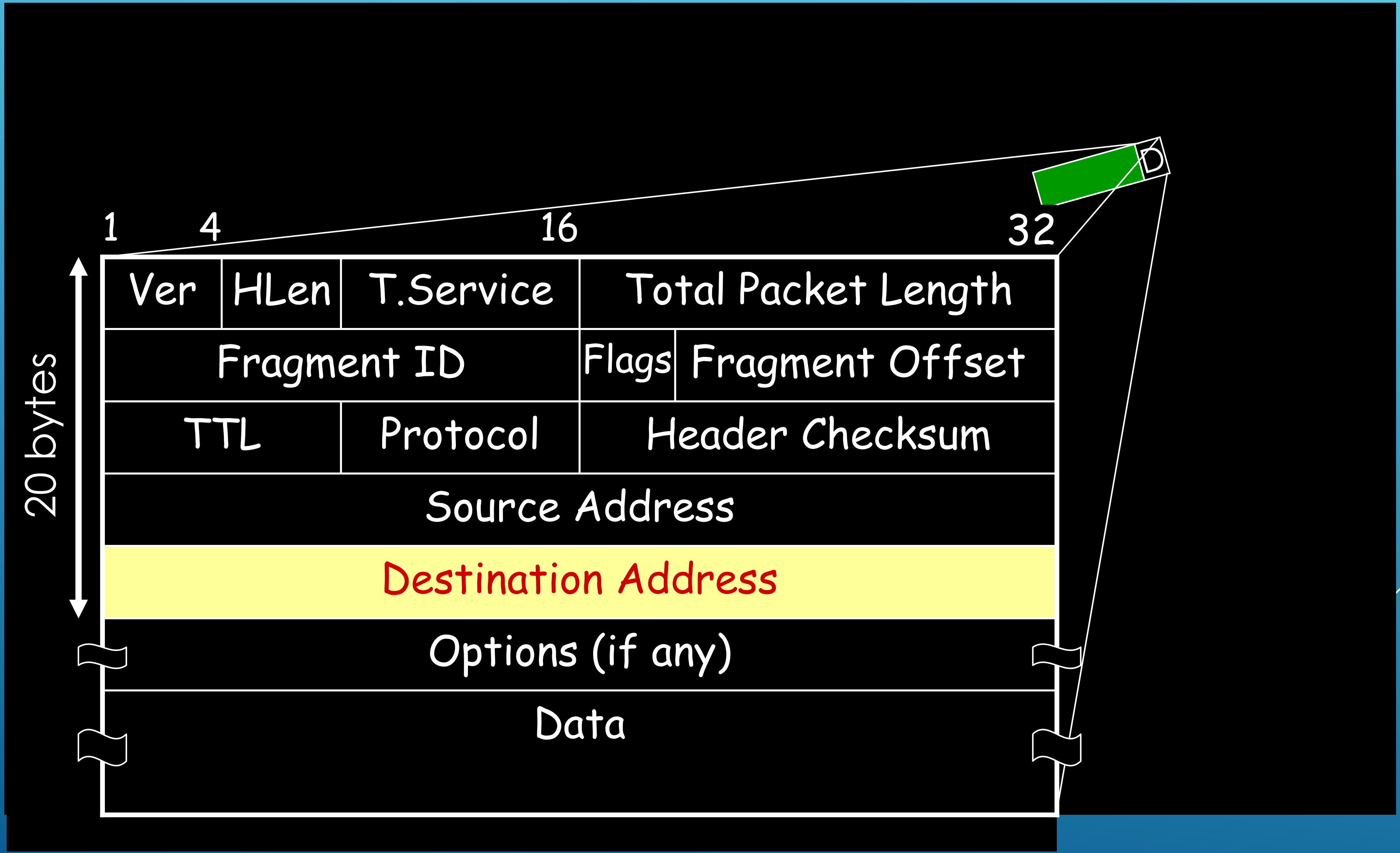
- Background
  - What is a router?
  - Why do we need faster routers?
  - Why are they hard to build?
- Architectures and techniques
  - The evolution of router architecture.
  - IP address lookup.
  - Packet buffering.
  - Switching.
- The Future

WHAT IS ~~ROUTING~~ FORWARDING?

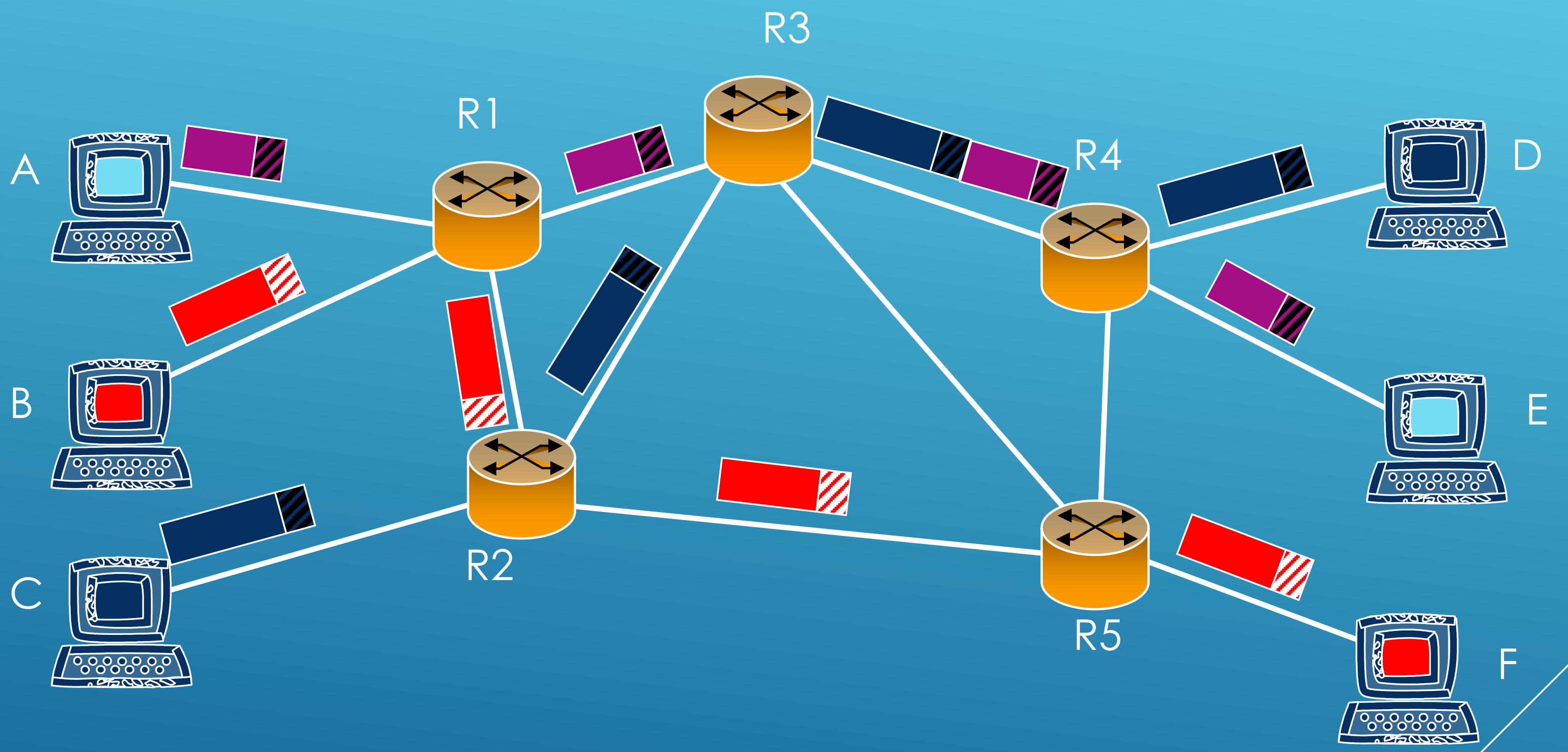


Destination	Next Hop
D	R3
E	R3
F	R5

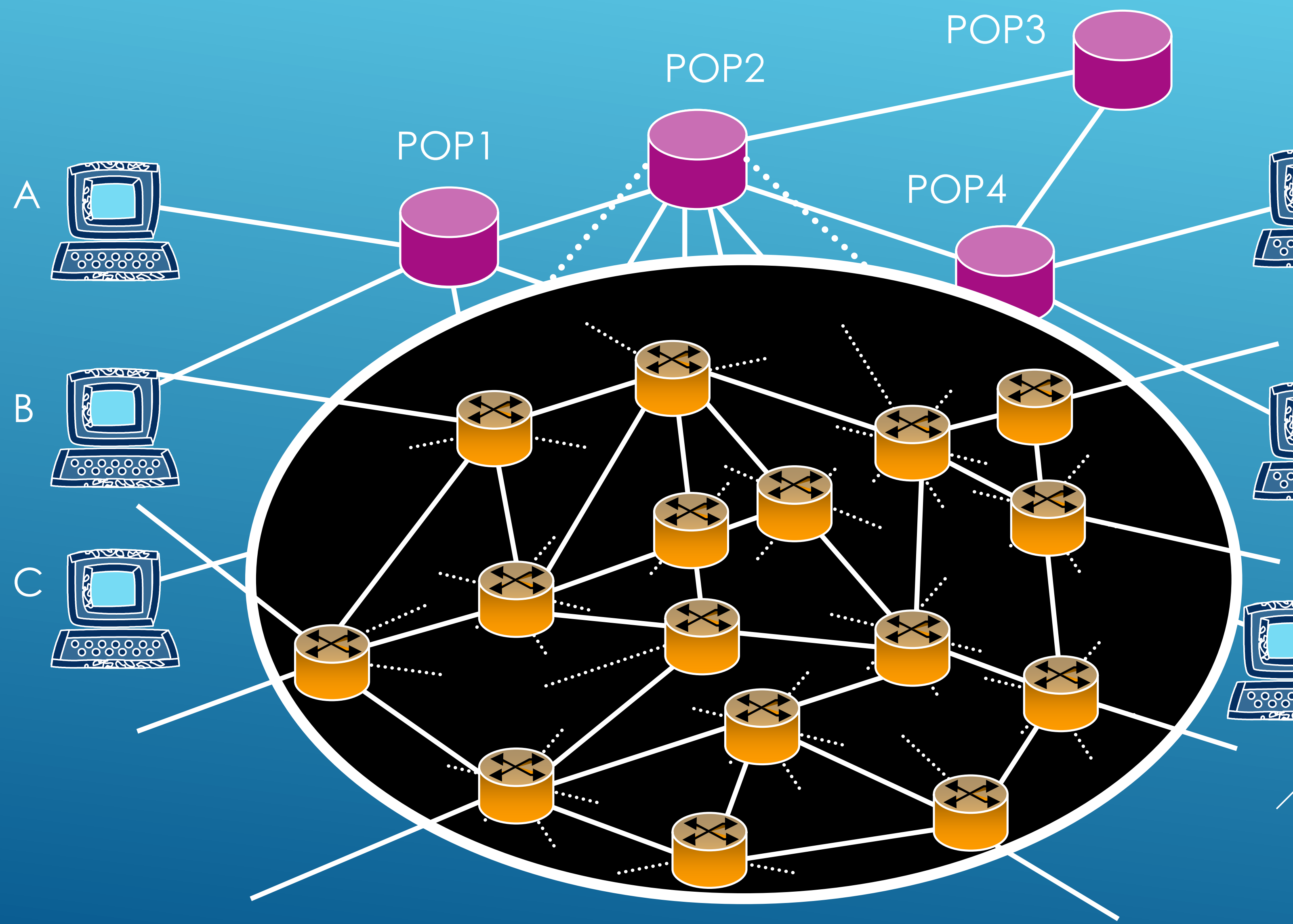
WHAT IS ROUTING?



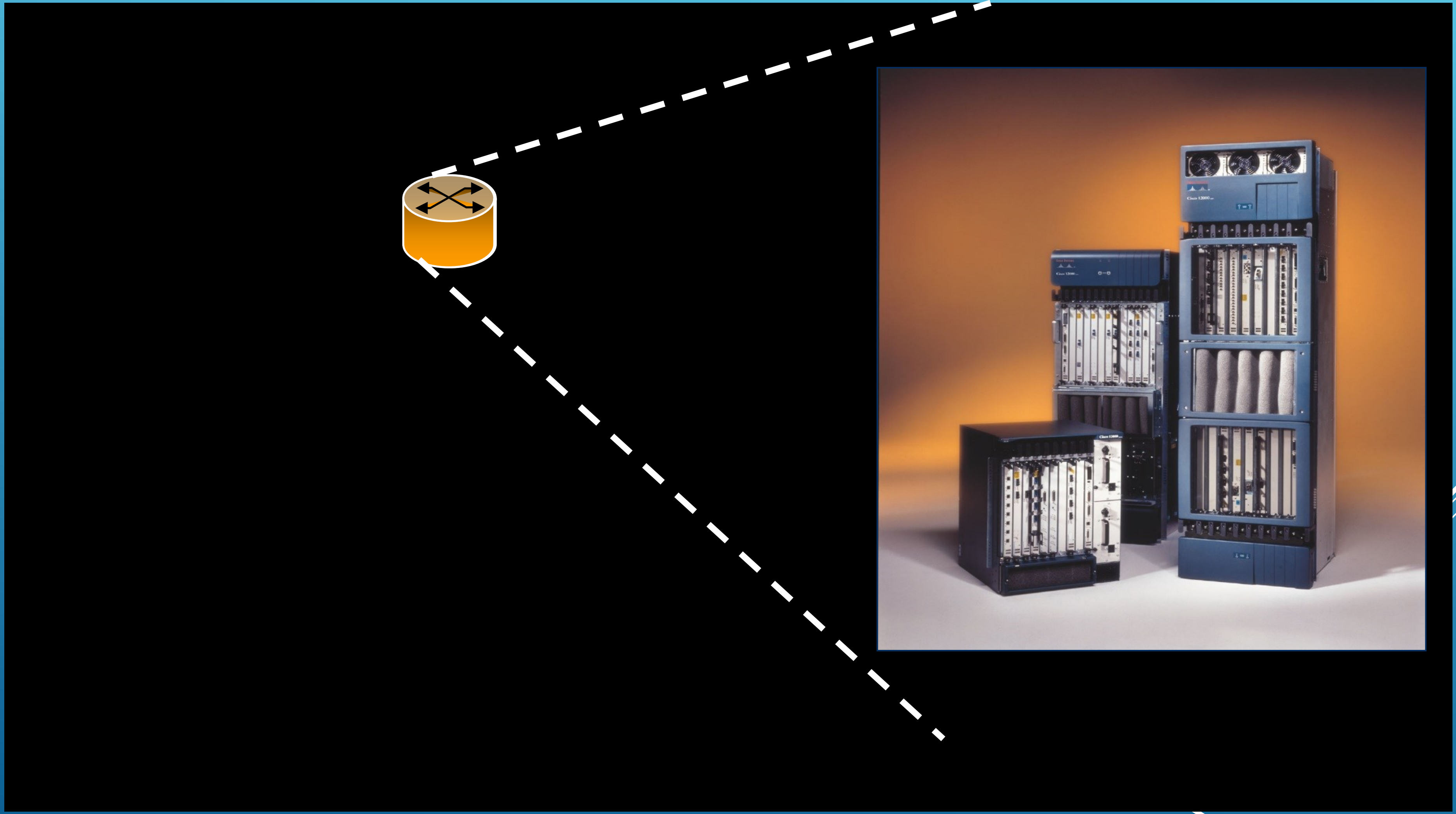
WHAT IS ROUTING?



POINTS OF PRESENCE (POPS)



WHERE HIGH PERFORMANCE ROUTERS ARE USED



WHAT A ROUTER LOOKS LIKE

Cisco GSR 12416

Capacity: 160Gb/s
Power: 4.2kW

Juniper M160

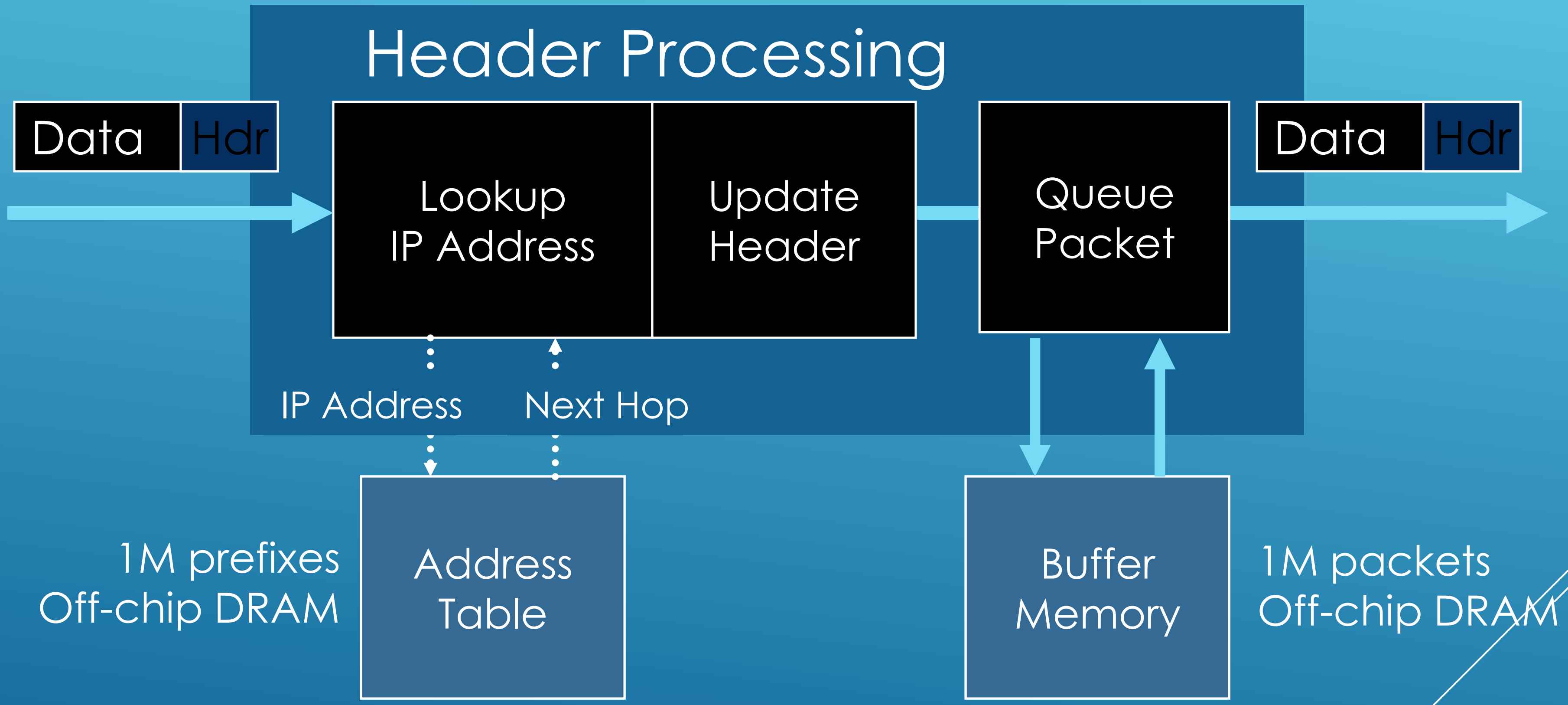
19”

3ft

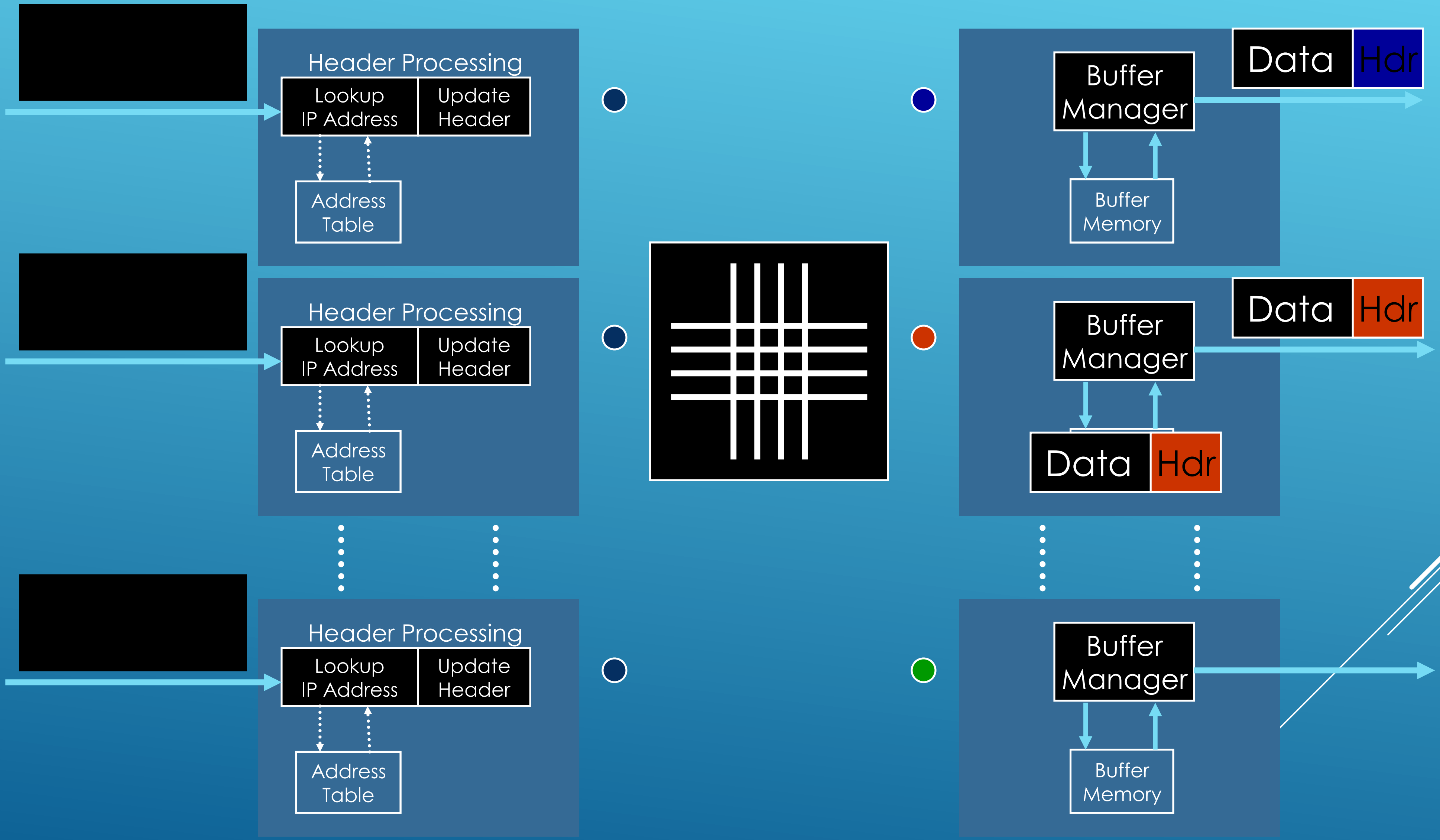
2.5ft

Capacity: 80Gb/s
Power: 2.6kW

GENERIC ROUTER ARCHITECTURE



GENERIC ROUTER ARCHITECTURE

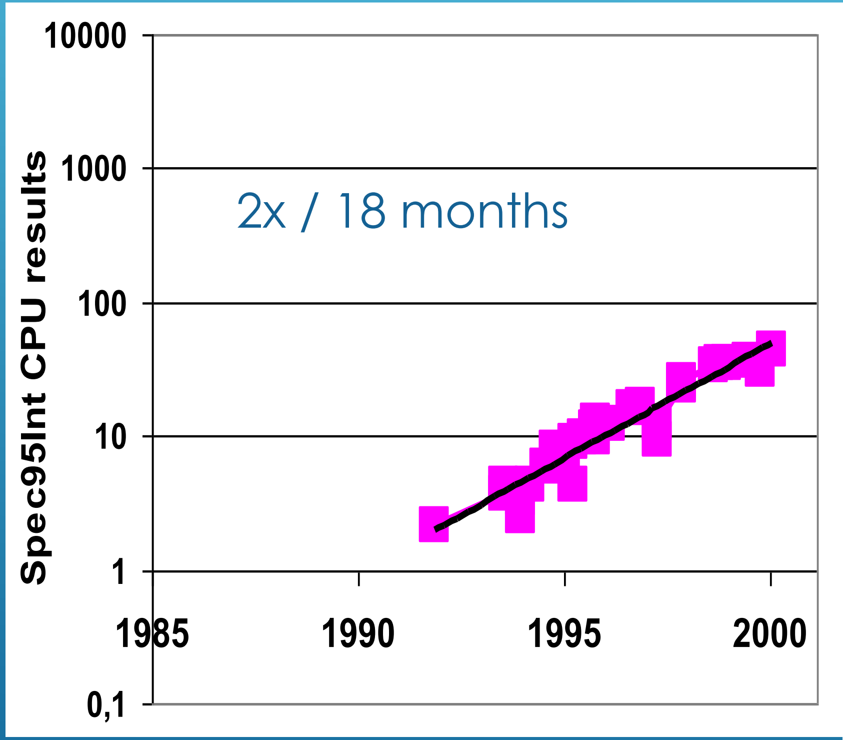


WHY DO WE NEED FASTER ROUTERS?

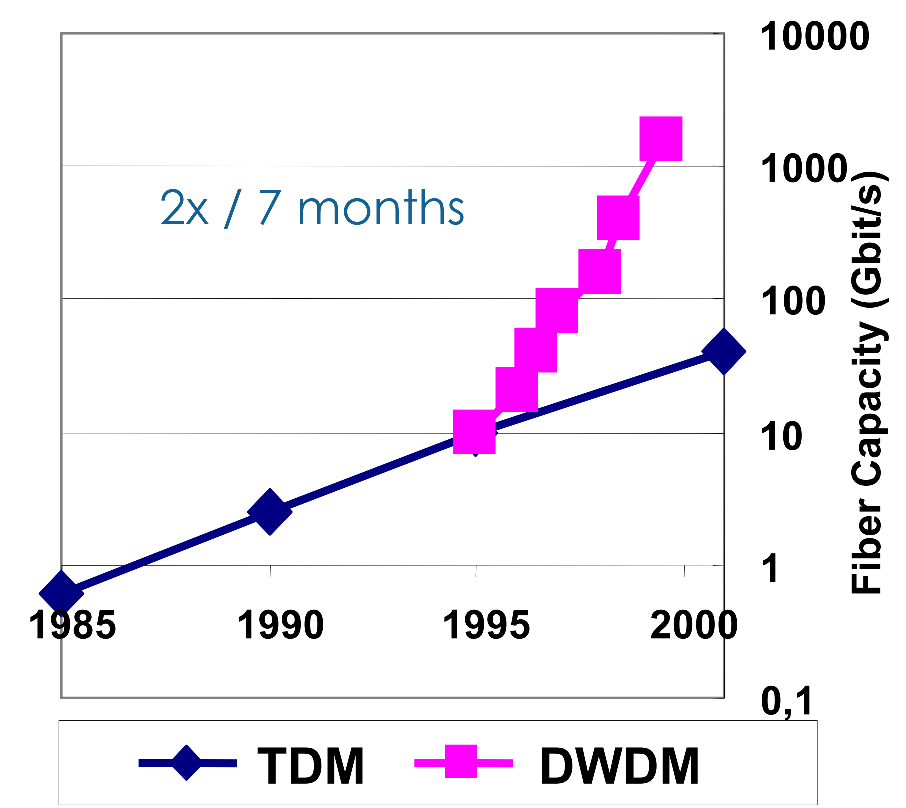
1. To prevent routers becoming the bottleneck in the Internet.
2. To increase POP capacity, and to reduce cost, size and power.

WHY WE NEED FASTER ROUTERS
1: *TO PREVENT ROUTERS FROM BEING THE BOTTLENECK*

Packet processing Power



Link Speed

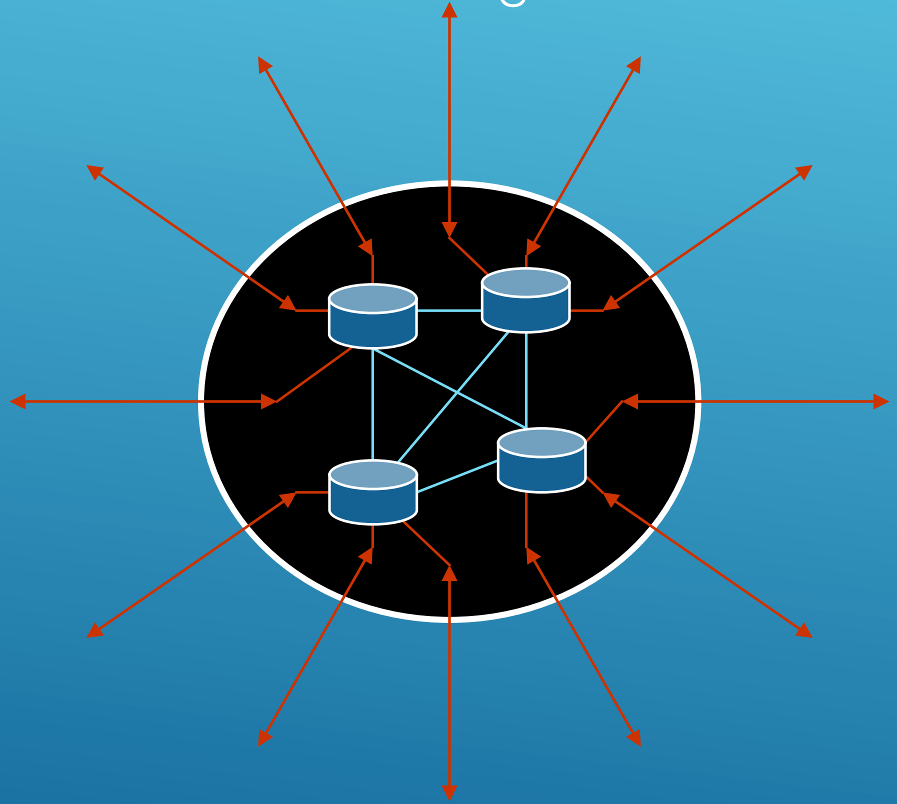


Source: SPEC95Int & David Miller, Stanford.

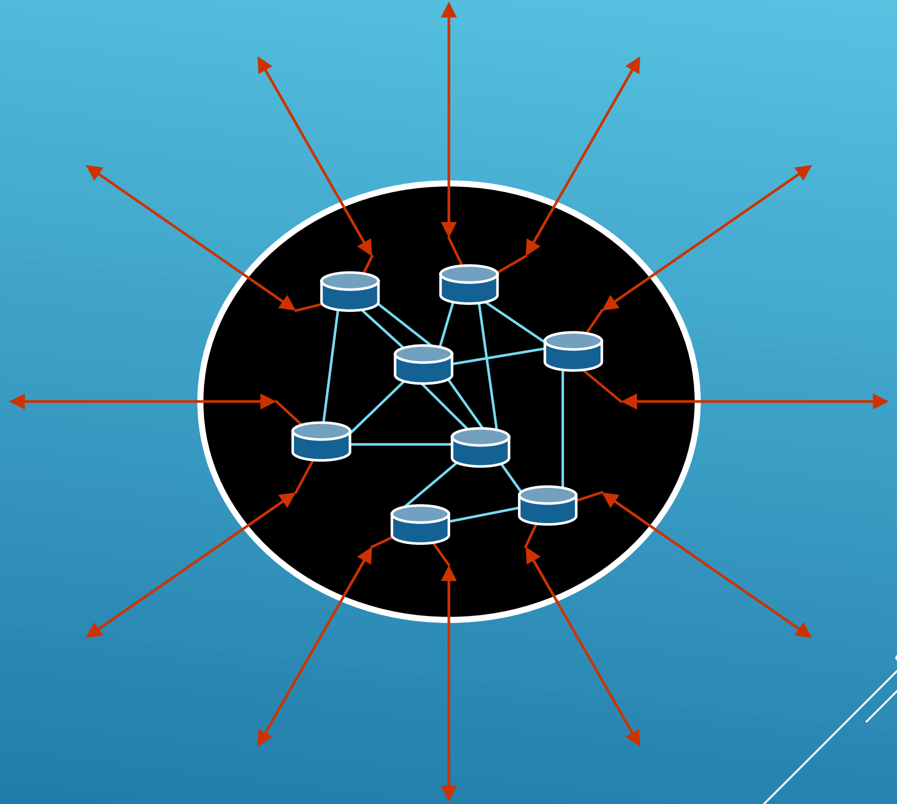
WHY WE NEED FASTER ROUTERS

2: TO REDUCE COST, POWER & COMPLEXITY OF POPS

POP with large routers



POP with smaller routers



- Ports: Price >$100k, Power > 400W.
- It is common for 50-60% of ports to be for interconnection.

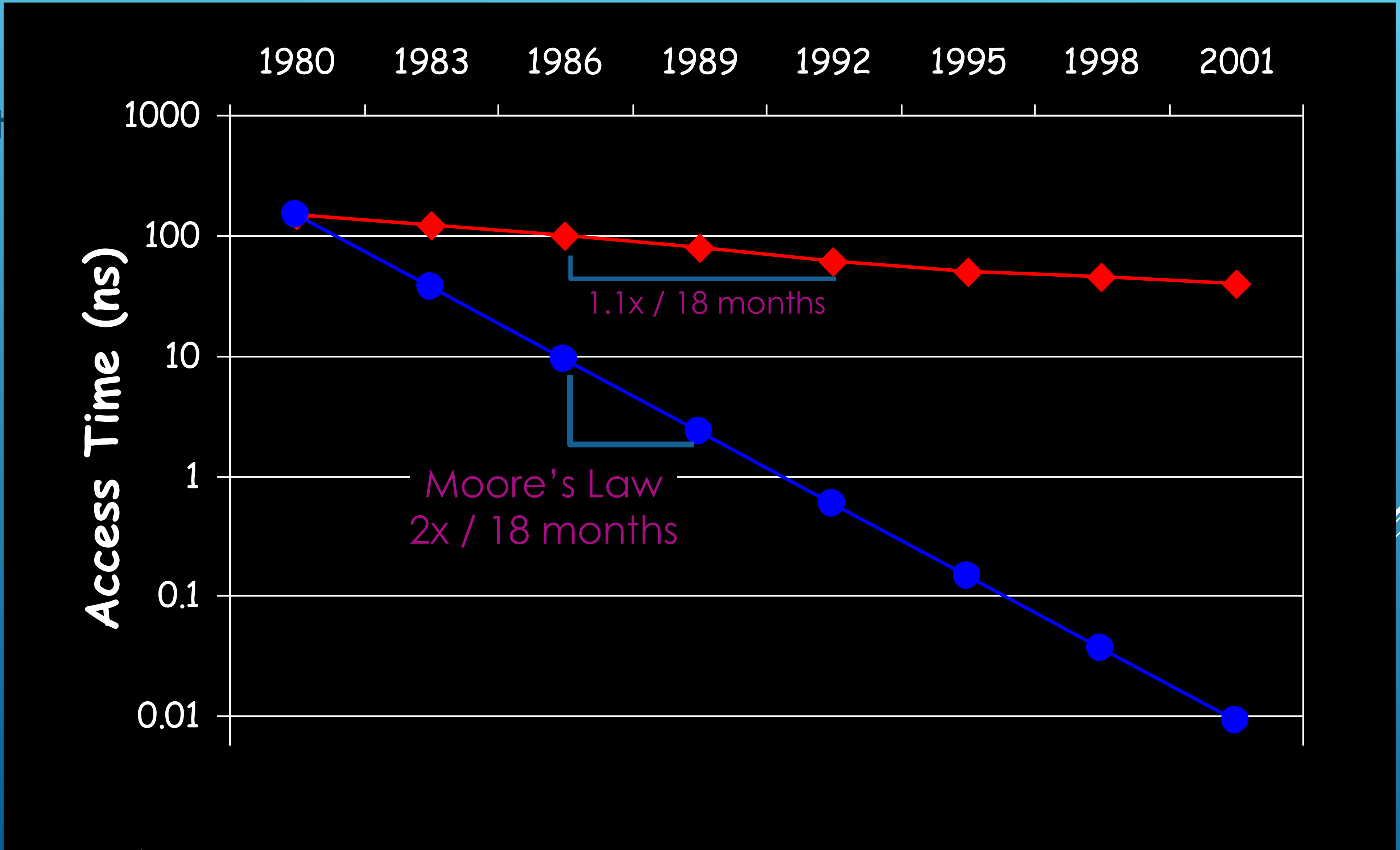
[redacted]

WHY ARE FAST ROUTERS DIFFICULT TO MAKE?

1. **It's hard to keep up with Moore's Law:**
    - The bottleneck is memory speed.
    - Memory speed is not keeping up with Moore's Law.

WHY ARE FAST ROUTERS DIFFICULT TO MAKE?

SPEED OF COMMERCIAL DRAM



WHY ARE FAST ROUTERS DIFFICULT TO MAKE?

1. **It's hard to keep up with Moore's Law:**
   - The bottleneck is memory speed.
   - Memory speed is not keeping up with Moore's Law.
2. **Moore's Law is too slow:**
   - Routers need to improve *faster* than Moore's Law.

ROUTER PERFORMANCE EXCEEDS MOORE'S LAW

Growth in capacity of commercial routers:

- Capacity 1992 ~ 2Gb/s
- Capacity 1995 ~ 10Gb/s
- Capacity 1998 ~ 40Gb/s
- Capacity 2001 ~ 160Gb/s
- Capacity 2003 ~ 640Gb/s

Average growth rate: 2.2x / 18 months.

OUTLINE

Background

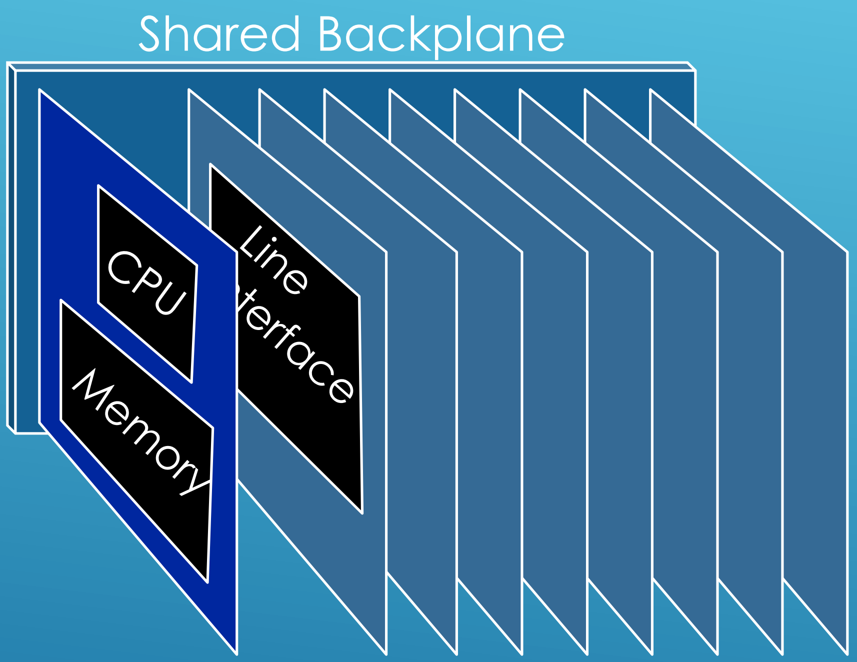
- What is a router?
- Why do we need faster routers?
- Why are they hard to build?

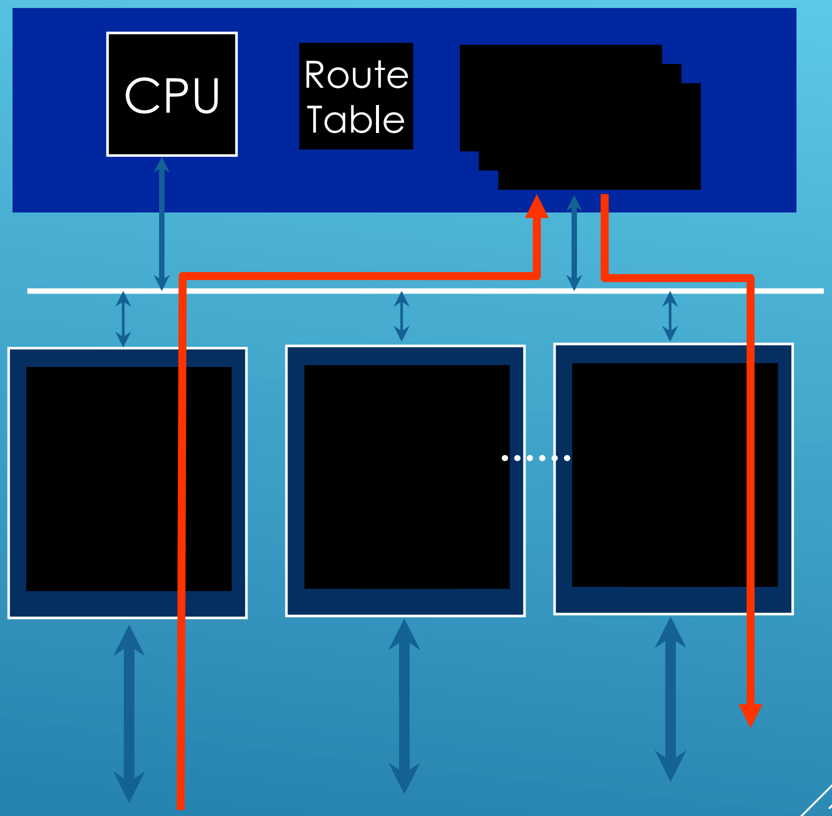
Architectures and techniques

- The evolution of router architecture.
- IP address lookup.
- Packet buffering.
- Switching.

The Future

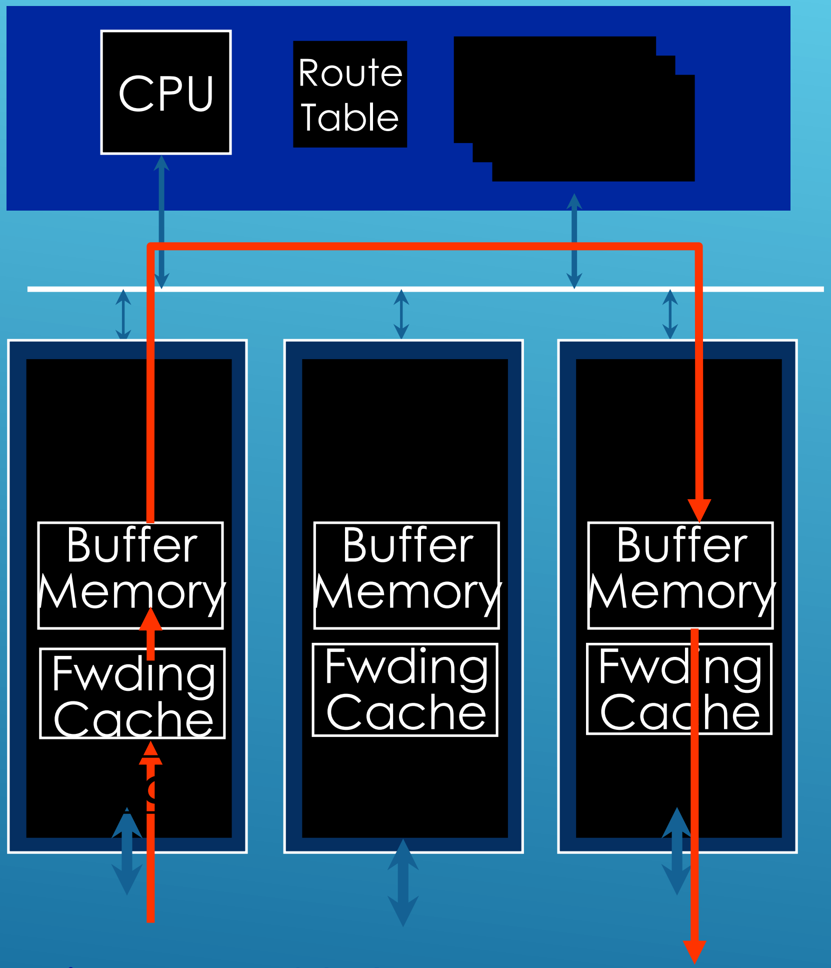
First Generation Routers





Typically <0.5Gb/s aggregate capacity

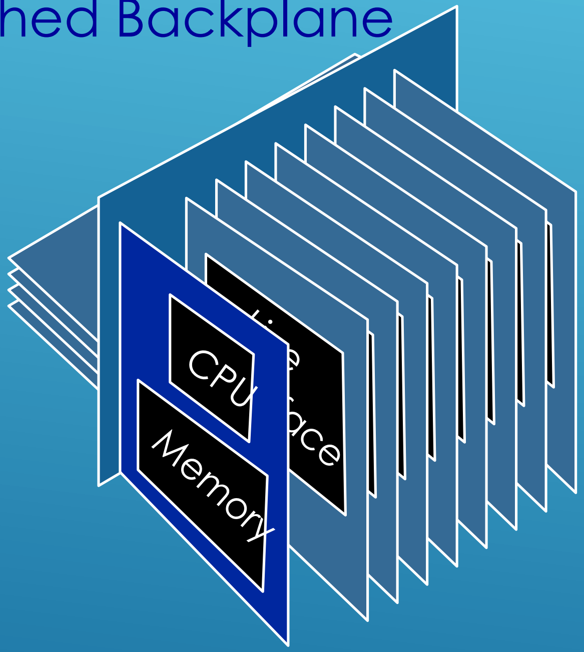
Second Generation Routers

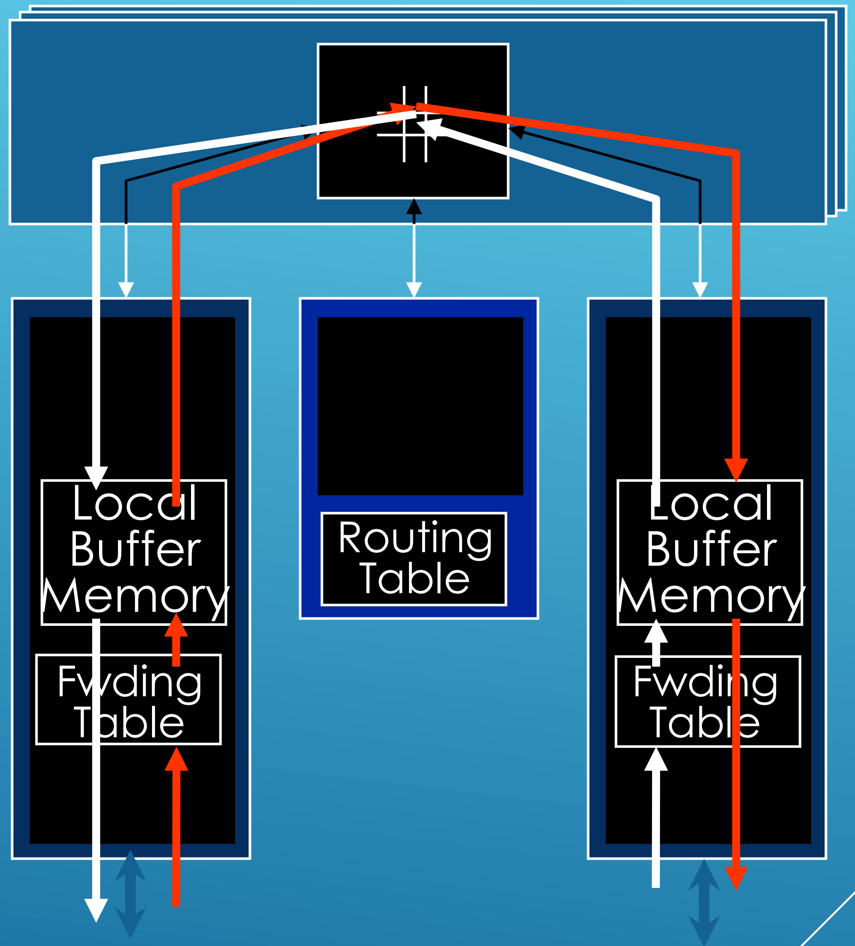


Typically <5Gb/s aggregate capacity

Third Generation Routers

Switched Backplane

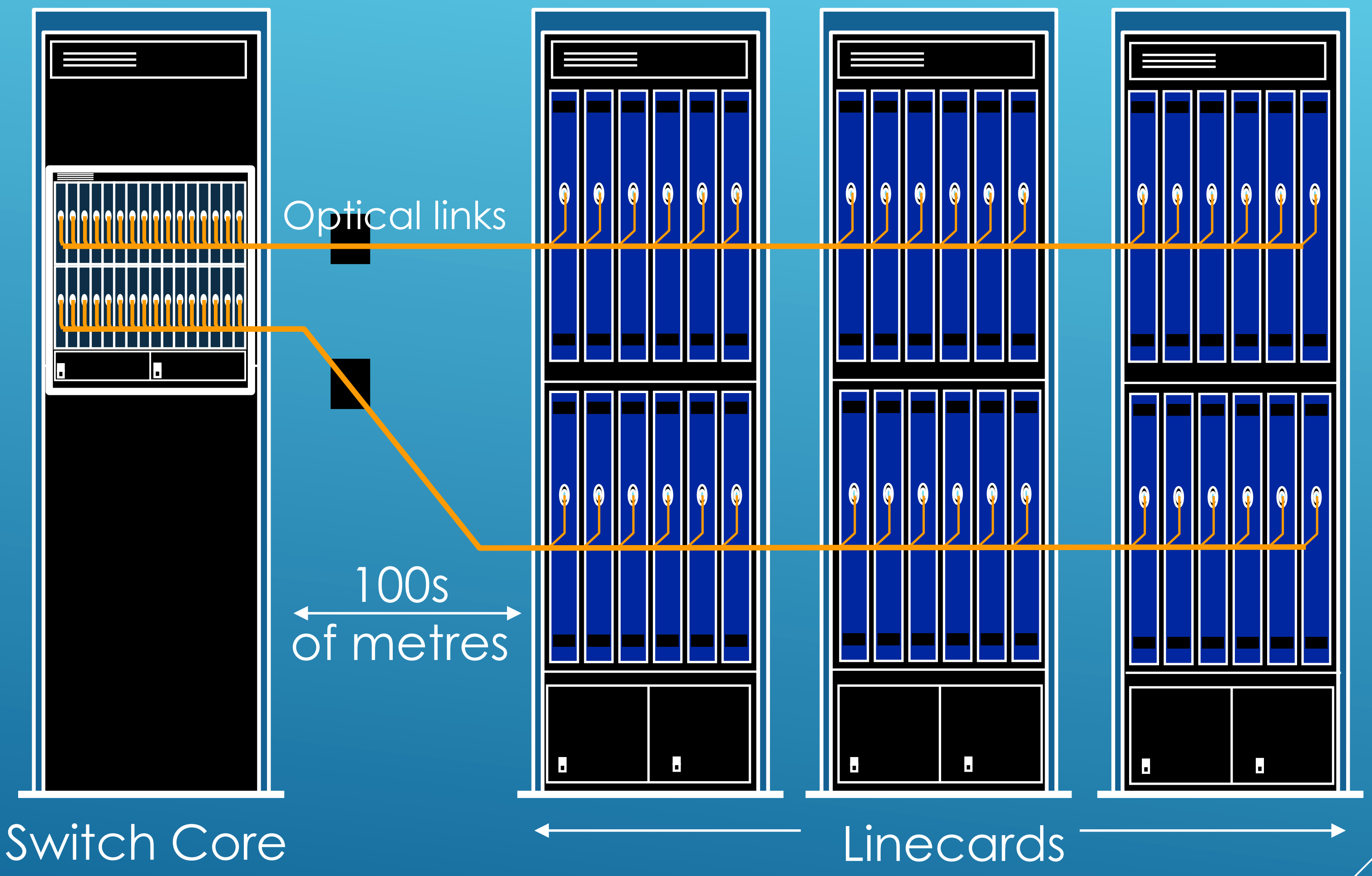




Typically <50Gb/s aggregate capacity

Fourth Generation Routers/Switches

Optics inside a router for the first time



0.3 - 10Tb/s routers in development

OUTLINE

Background

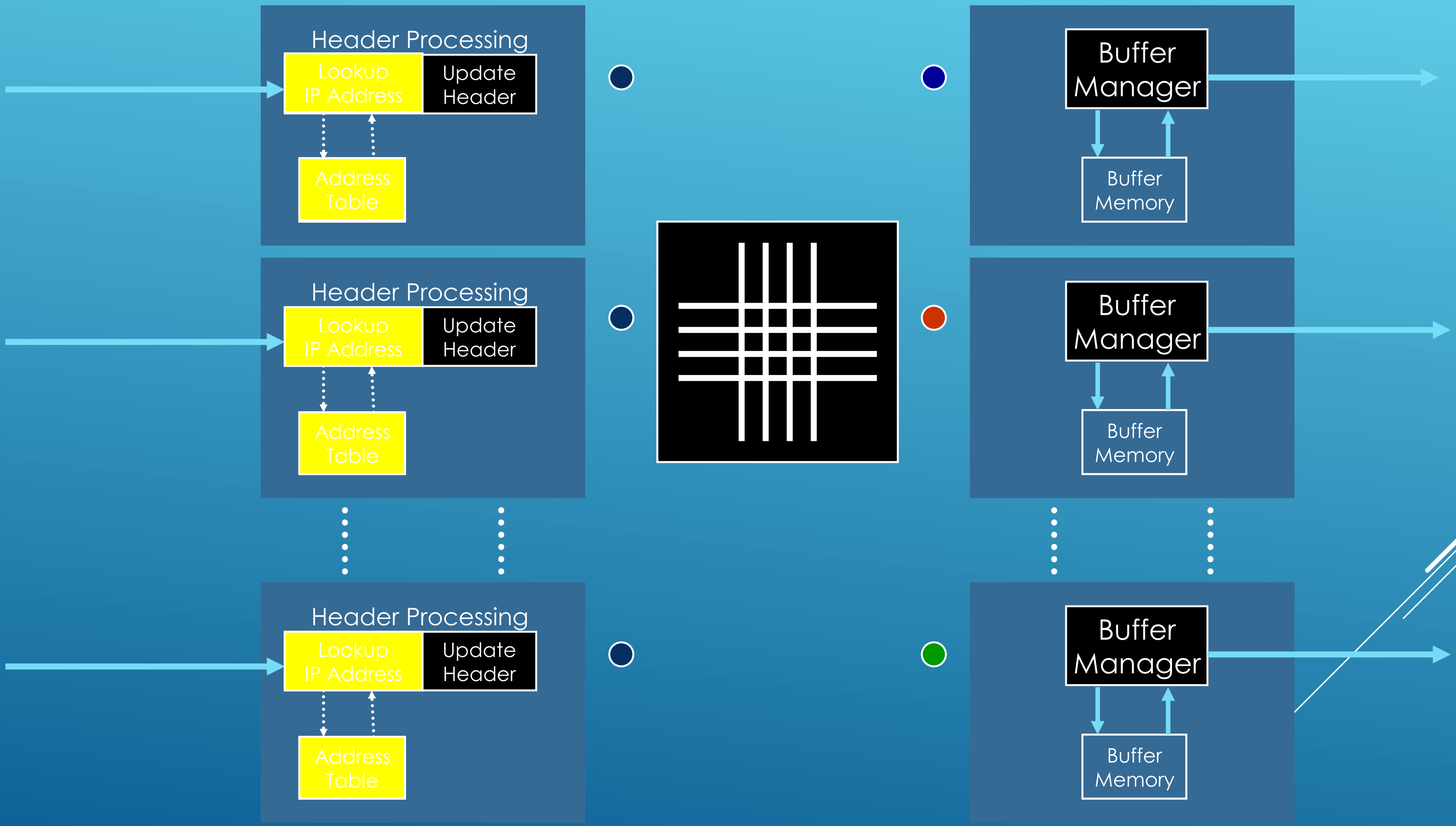
- What is a router?
- Why do we need faster routers?
- Why are they hard to build?

Architectures and techniques

- The evolution of router architecture.
- IP address lookup.
- Packet buffering.
- Switching.

The Future

GENERIC ROUTER ARCHITECTURE

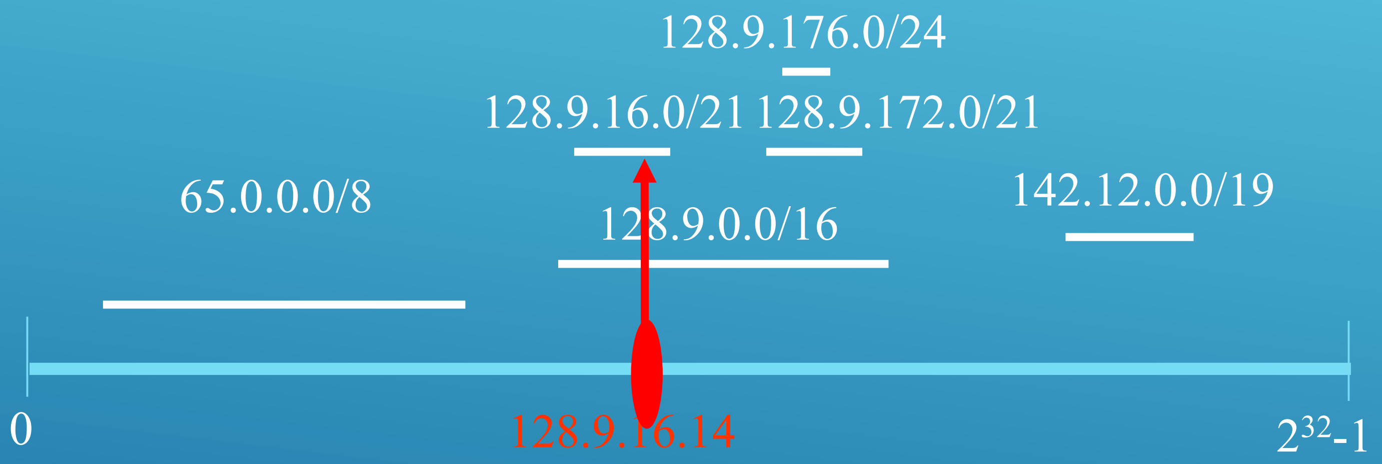


IP ADDRESS LOOKUP

Why it's thought to be hard:

1. It's not an exact match: it's a longest prefix match.
2. The table is large: about 120,000 entries today, and growing.
3. The lookup must be fast: about 30ns for a 10Gb/s line.

IP LOOKUPS FIND LONGEST PREFIXES



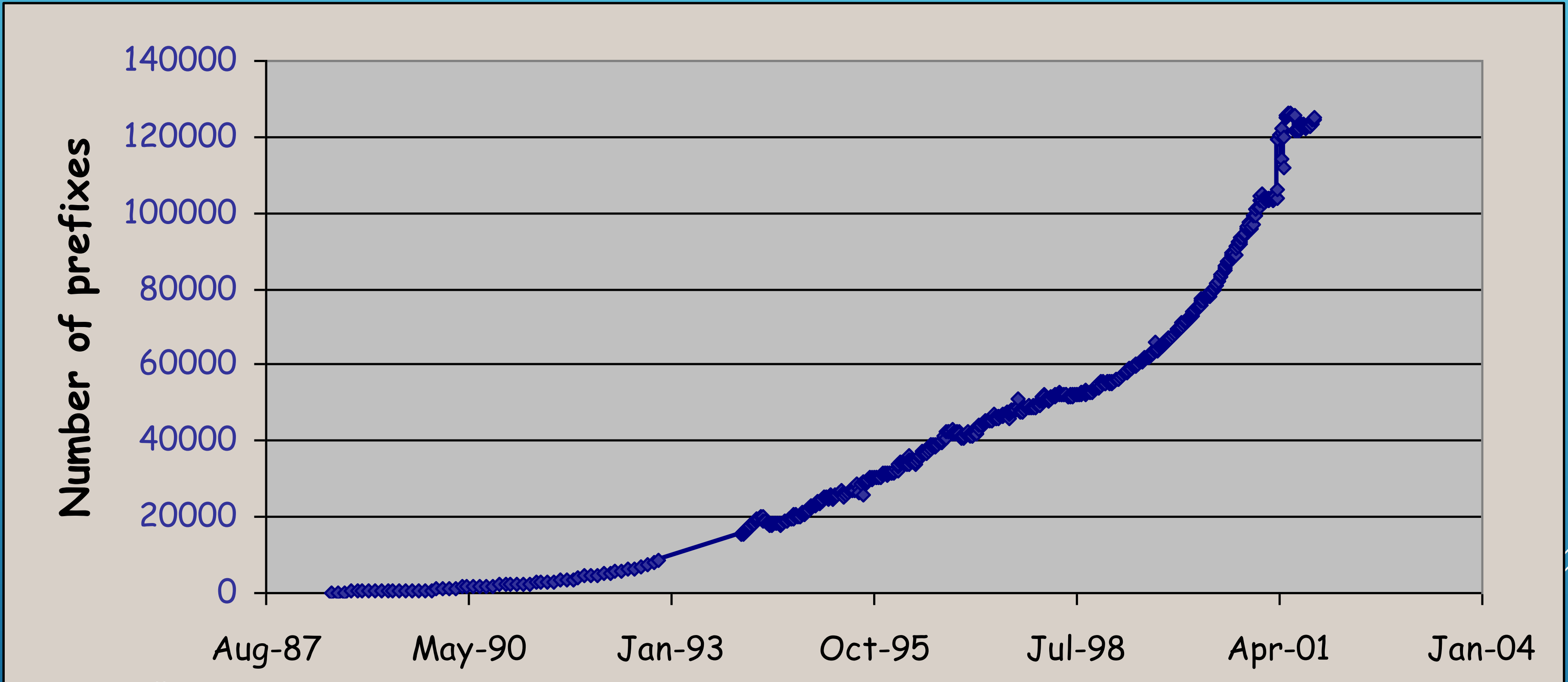
Routing lookup: Find the **longest** matching prefix (aka the most specific route) among all prefixes that match the destination address.

IP ADDRESS LOOKUP

Why it's thought to be hard:

1. It's not an exact match: it's a longest prefix match.
2. The table is large: about 120,000 entries today, and growing.
3. The lookup must be fast: about 30ns for a 10Gb/s line.

ADDRESS TABLES ARE LARGE



Source: Geoff Huston, Oct 2001

IP ADDRESS LOOKUP

Why it’s thought to be hard:

1. It’s not an exact match: it’s a longest prefix match.
2. The table is large: about 120,000 entries today, and growing.
3. The lookup must be fast: about 30ns for a 10Gb/s line.

LOOKUPS MUST BE FAST

Year	Line	40B packets (Mpkt/s)
2001	10Gb/s	31.25

OUTLINE

Background

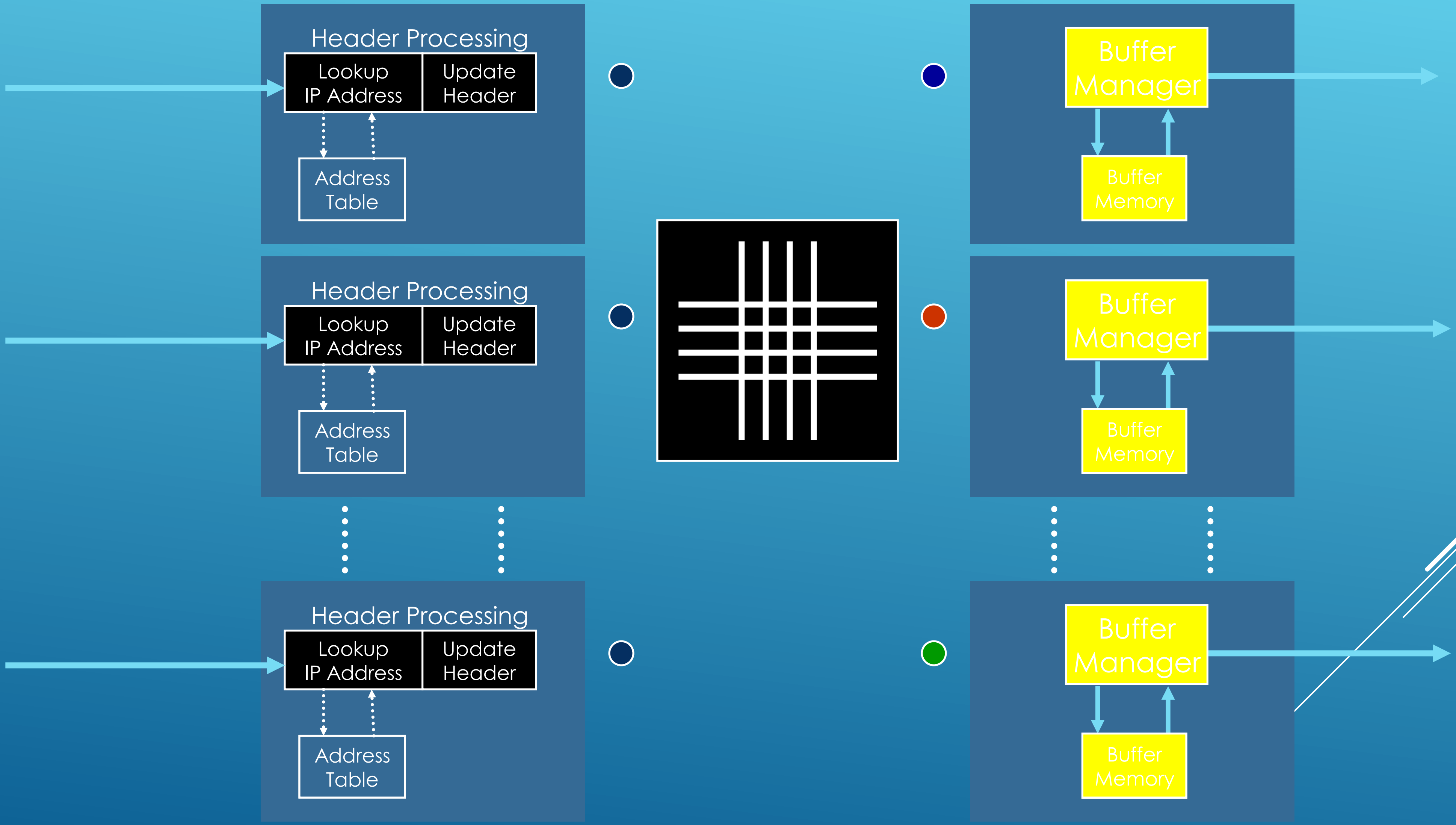
- What is a router?
- Why do we need faster routers?
- Why are they hard to build?

Architectures and techniques

- The evolution of router architecture.
- IP address lookup.
- Packet buffering.
- Switching.

The Future

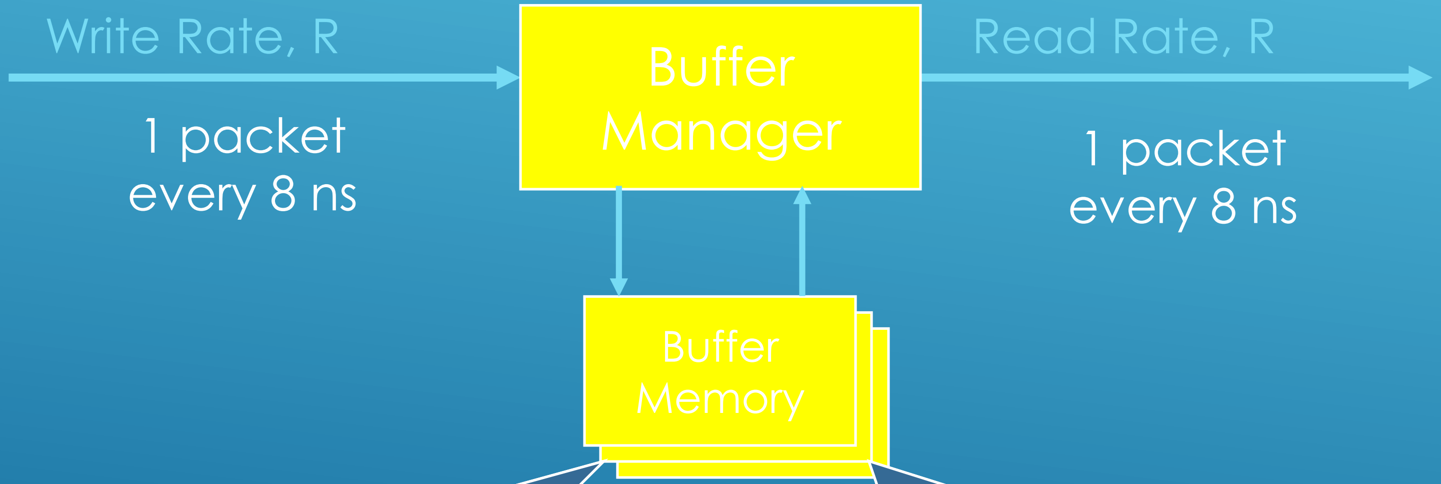
GENERIC ROUTER ARCHITECTURE



FAST PACKET BUFFERS

Example: 40Gb/s packet buffer

Size = RTT*BW = 10Gb; 40 byte packets



Use SRAM?

\+ fast enough random access time, but

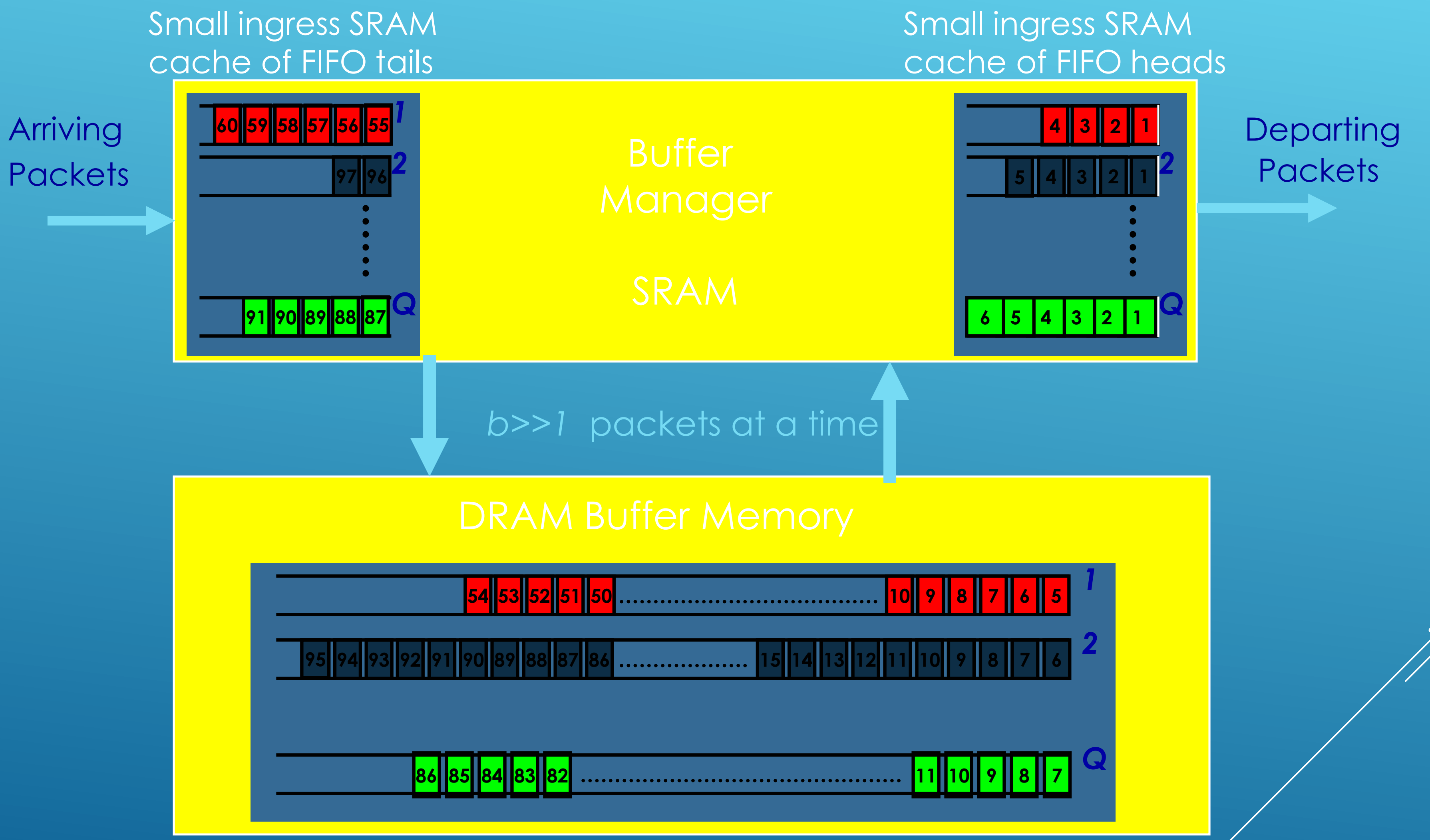
\- too low density to store 10Gb of data.

Use DRAM?

\+ high density means we can store data, but

\- too slow (50ns random access time).

PACKET CACHES



OUTLINE

Background

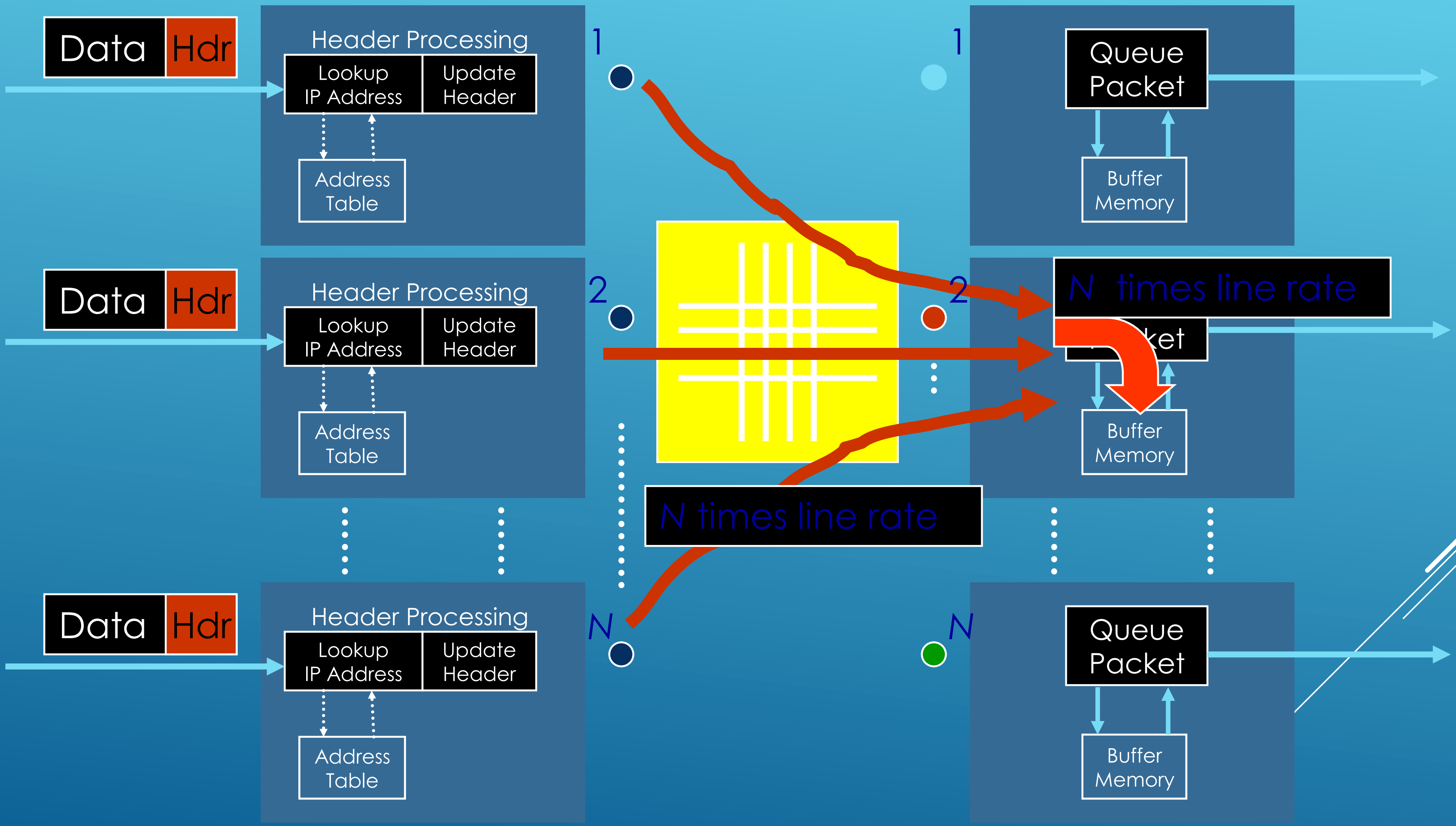
- What is a router?
- Why do we need faster routers?
- Why are they hard to build?

Architectures and techniques

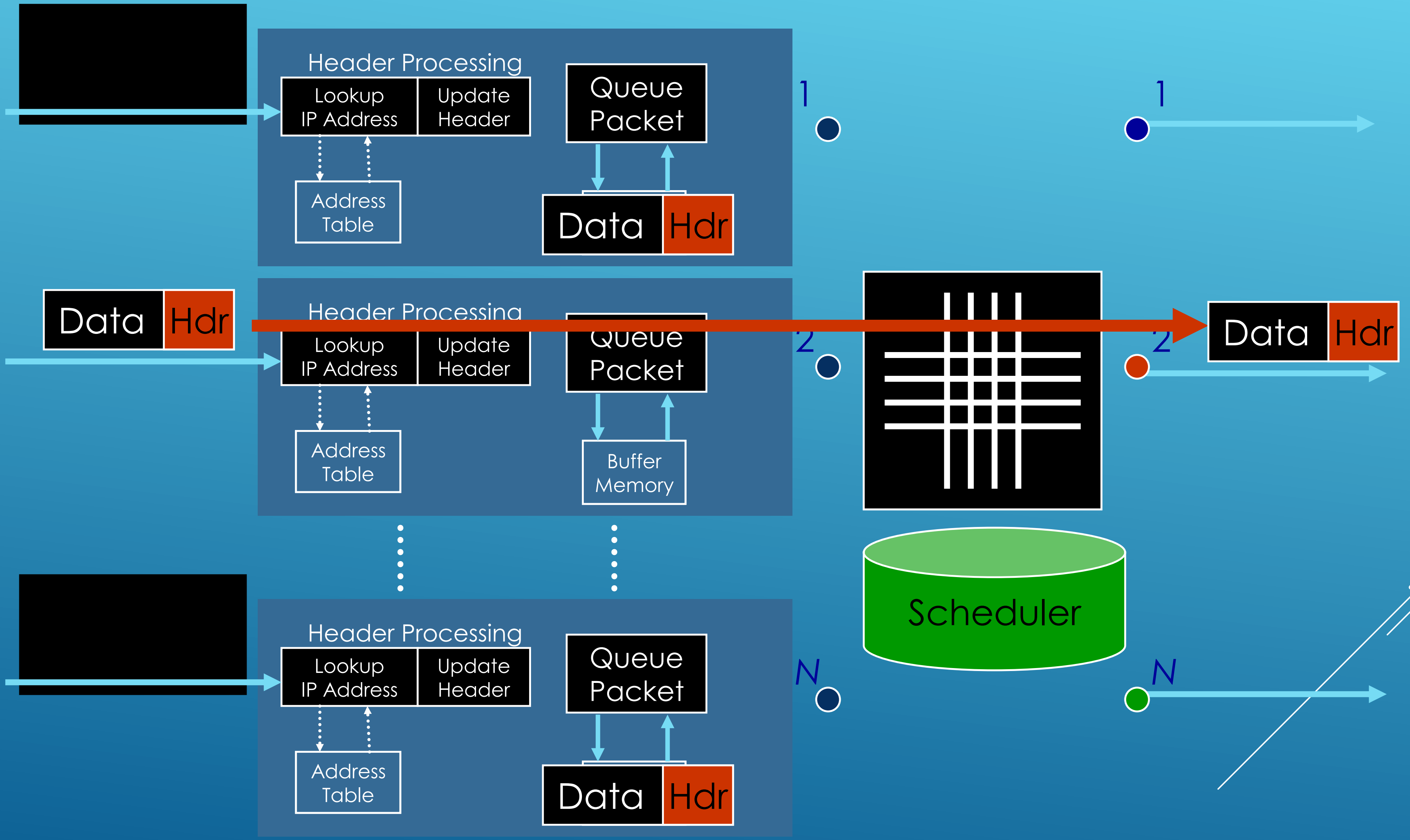
- The evolution of router architecture.
- IP address lookup.
- Packet buffering.
- Switching.

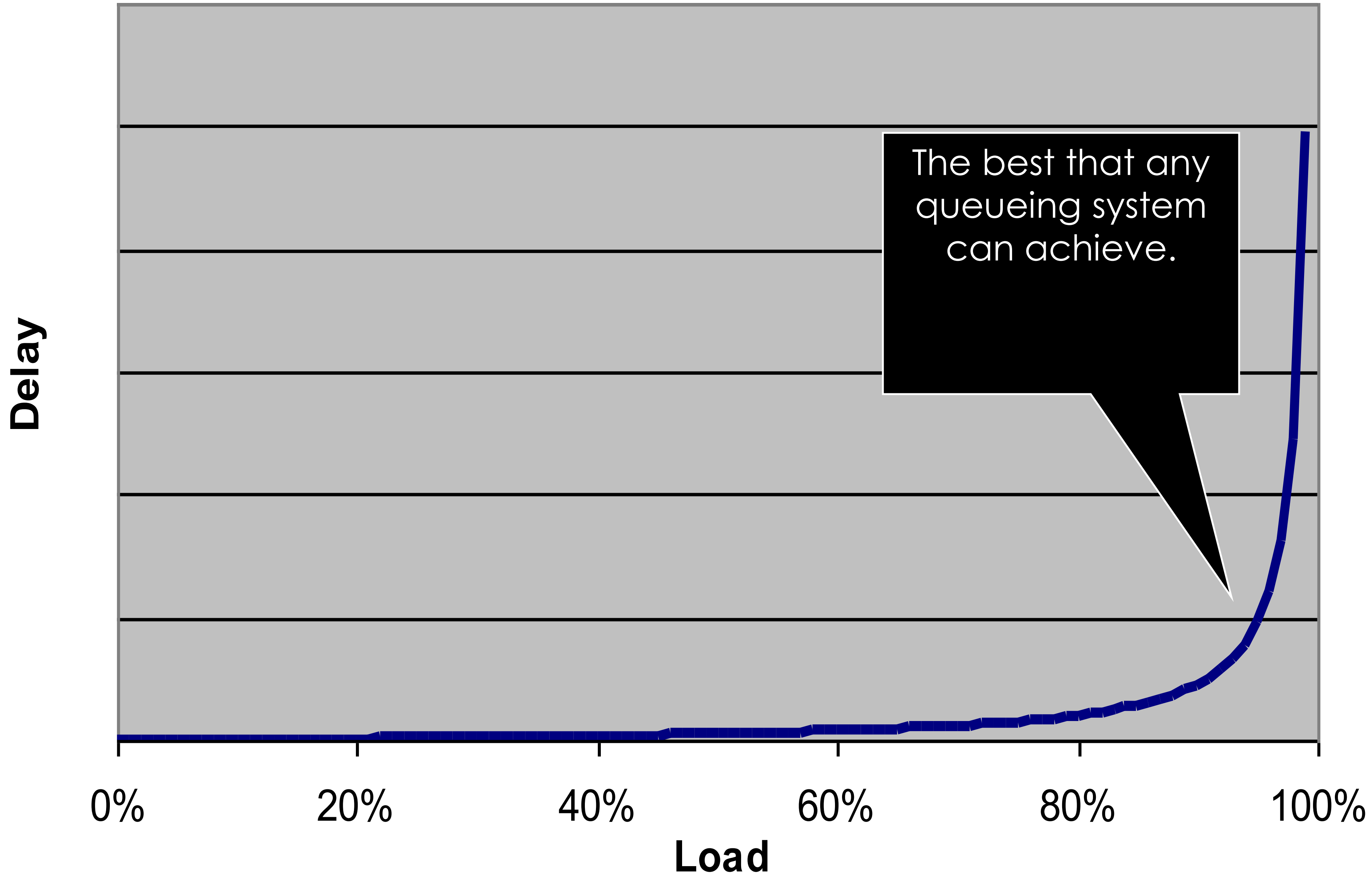
The Future

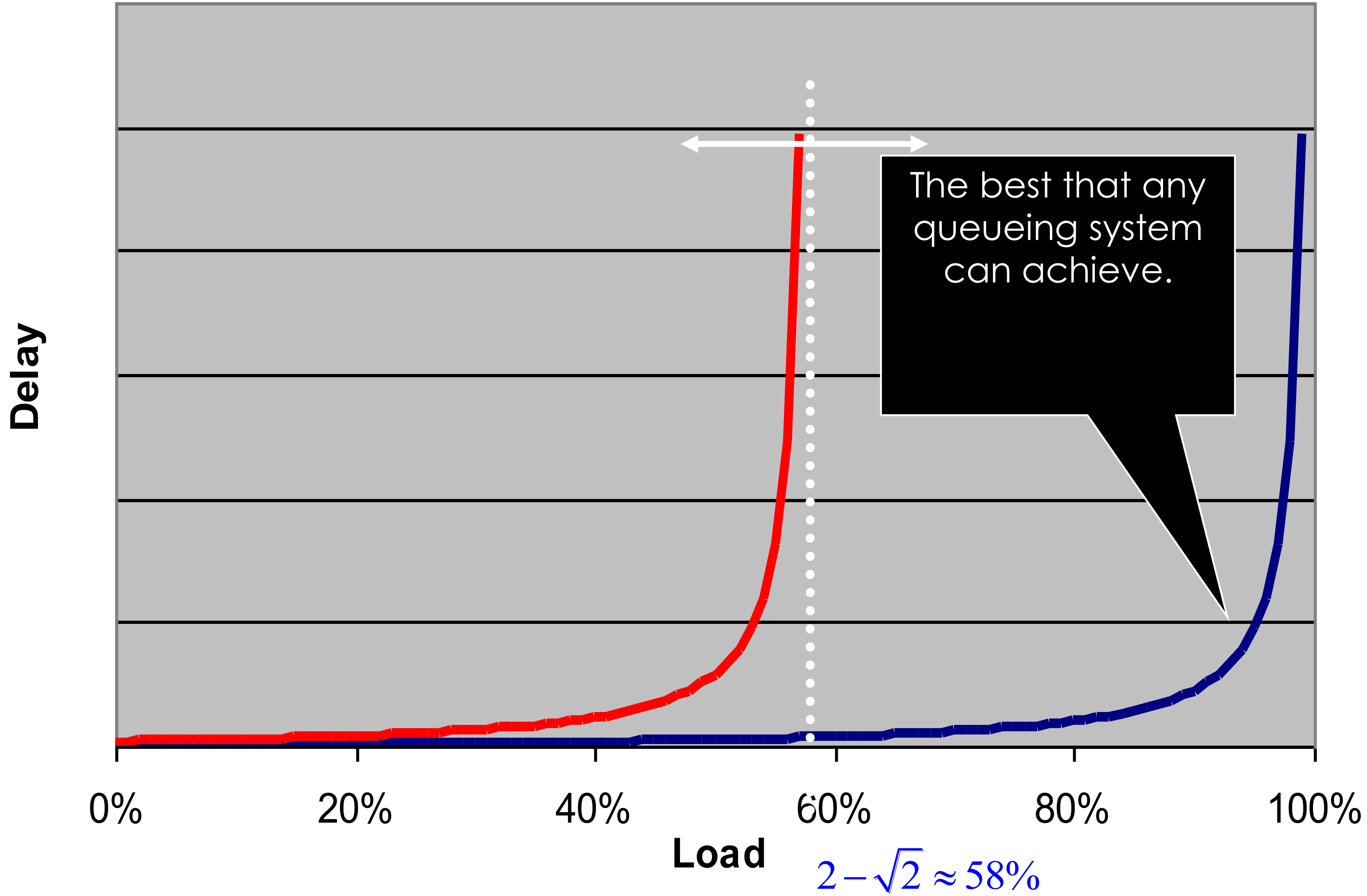
GENERIC ROUTER ARCHITECTURE



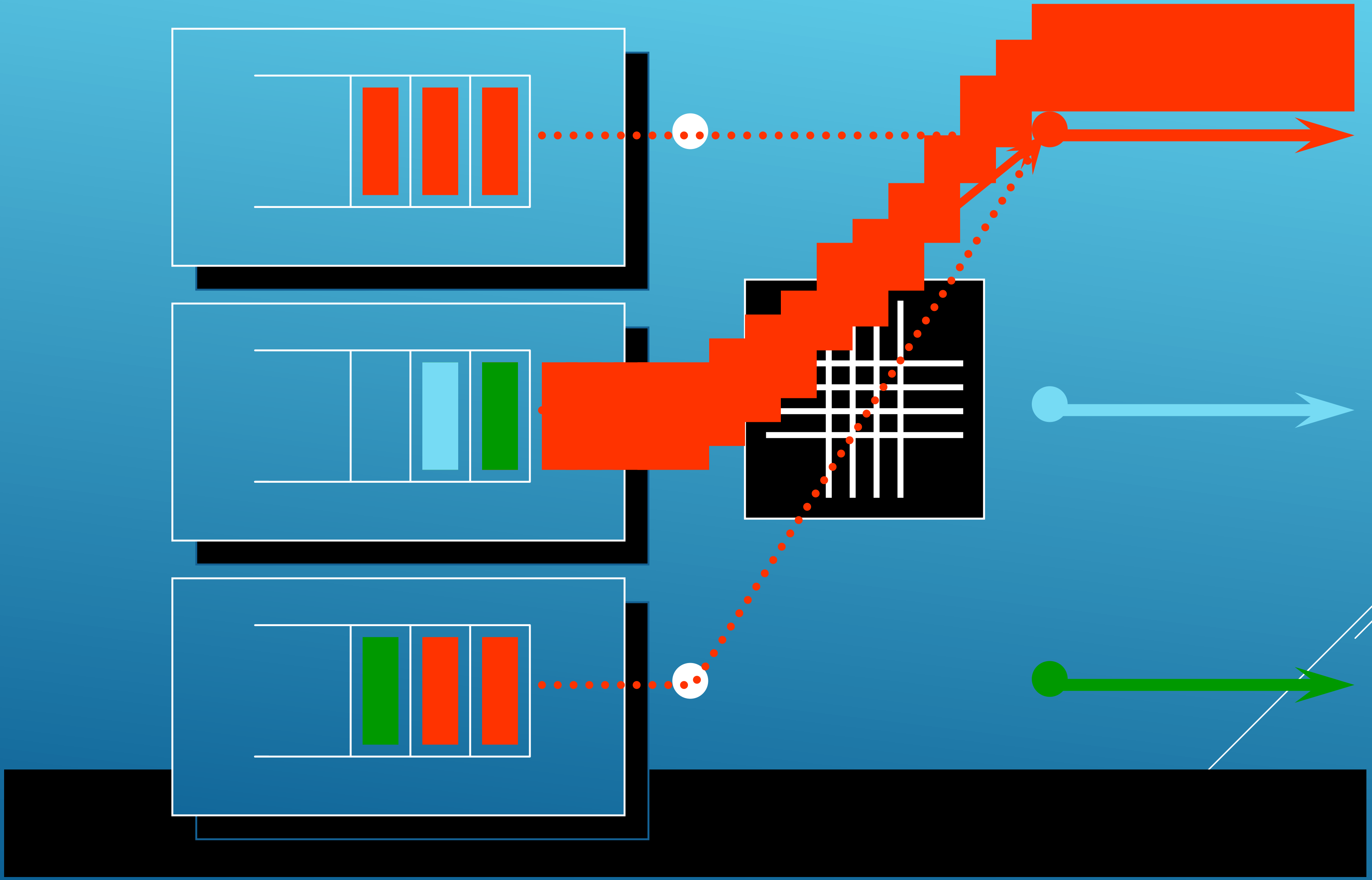
GENERIC ROUTER ARCHITECTURE



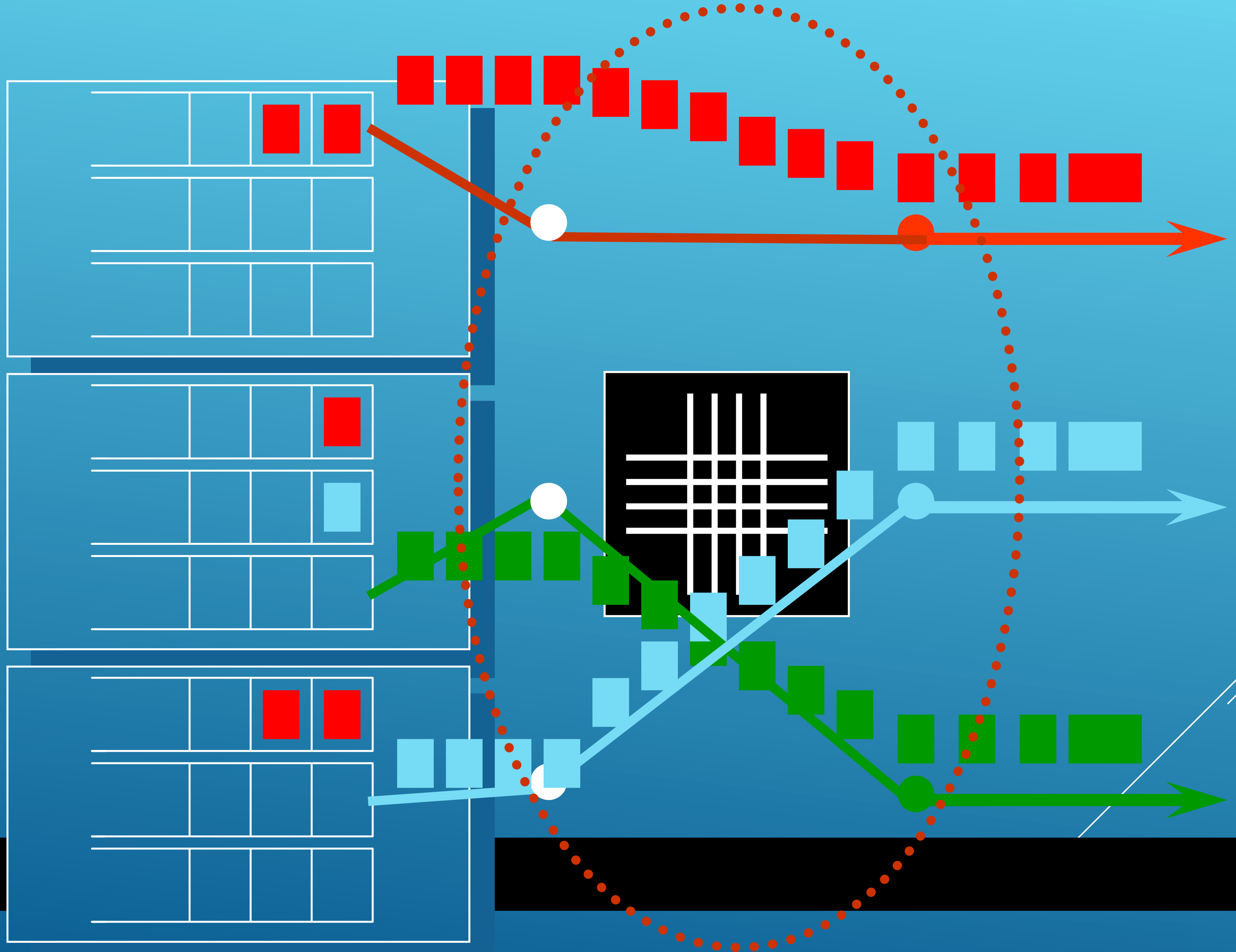

The best that any queueing system can achieve.
Delay
0%
20%
40%
60%
80%
100%
Load

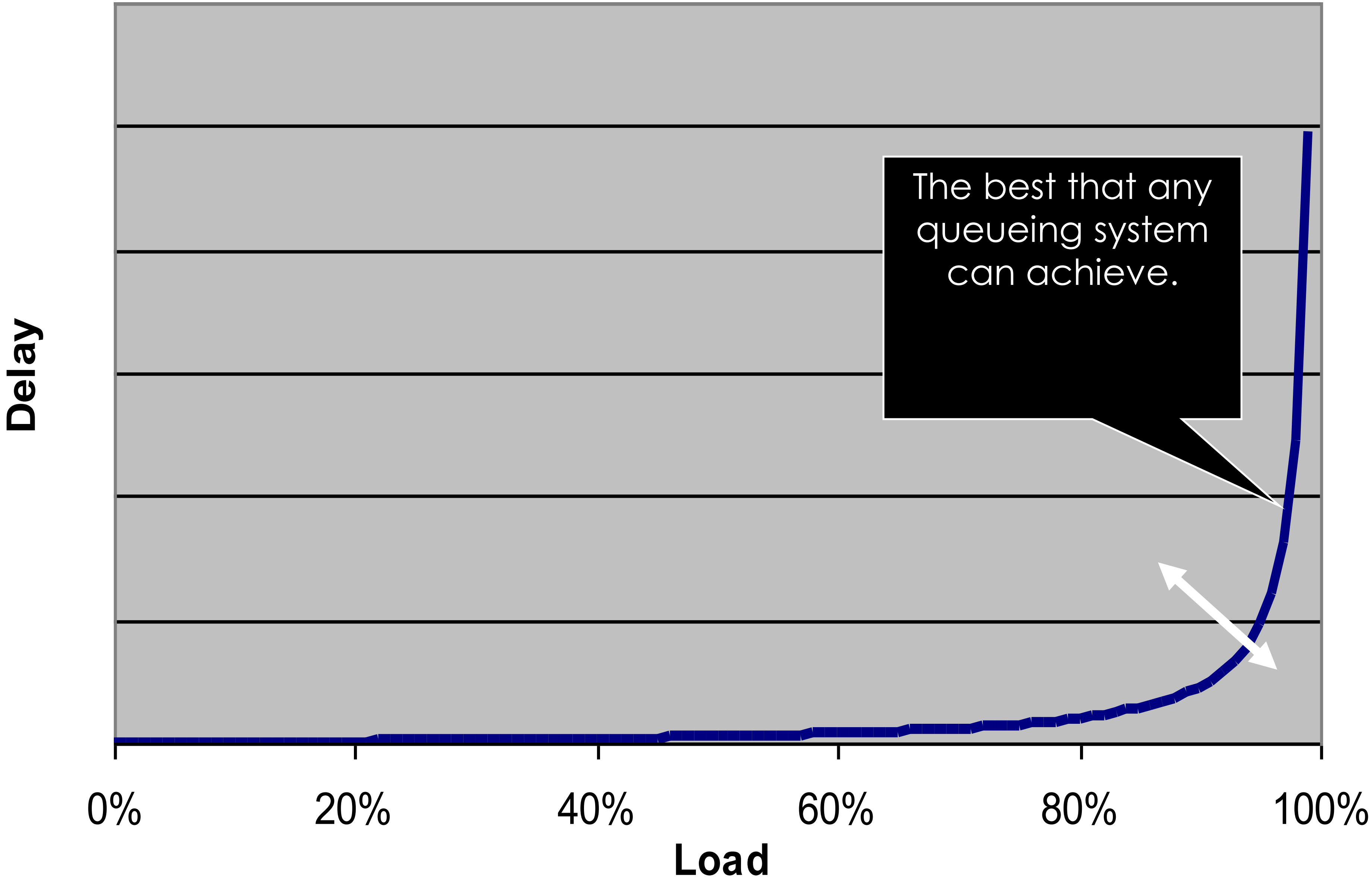

The best that any queueing system can achieve.
Delay
0%
20%
40%
60%
80%
100%
Load
$2-\sqrt{2} \approx 58\%$

HEAD OF LINE BLOCKING

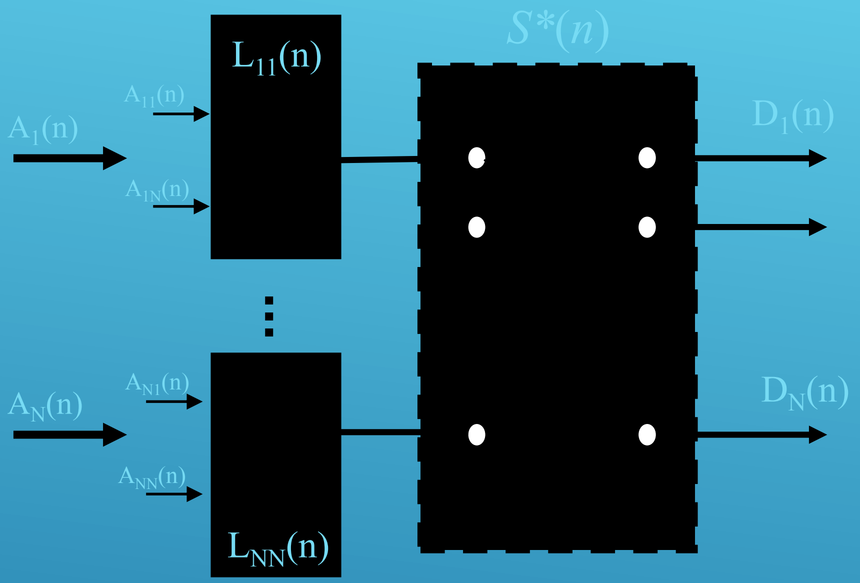


VIRTUAL OUTPUT QUEUES




Delay
The best that any queueing system can achieve.
0%
20%
40%
60%
80%
100%
Load

MAXIMUM WEIGHT MATCHING



$$S^*(n) = \arg\max_{\underline{S}(n)} (\underline{L}^T(n) \cdot \underline{S}(n))$$

“Request” Graph

Bipartite Match

OUTLINE

Background

- What is a router?
- Why do we need faster routers?
- Why are they hard to build?

Architectures and techniques

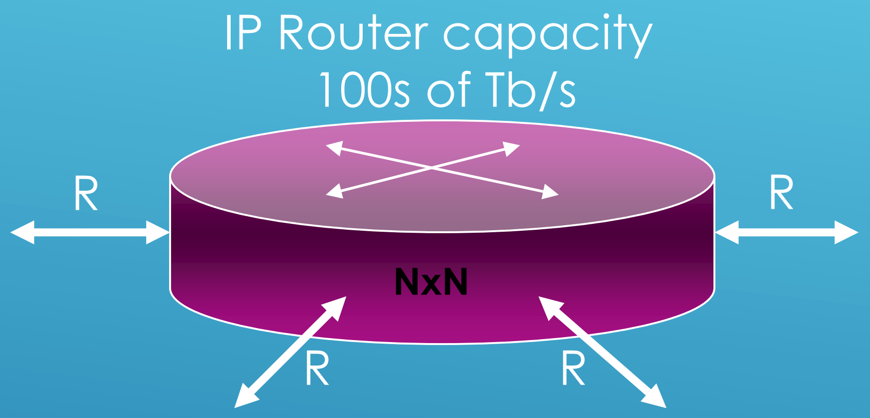
- The evolution of router architecture.
- IP address lookup.
- Packet buffering.
- Switching.

The Future

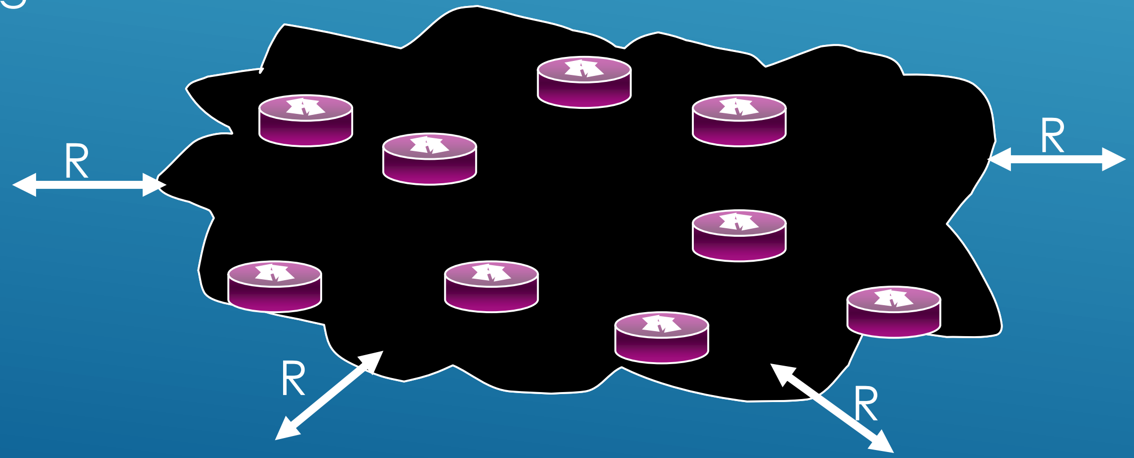
- More parallelism.
- Eliminating schedulers.
- Introducing optics into routers.
- Natural evolution to circuit switching?

EXTERNAL PARALLELISM:
MULTIPLE PARALLEL ROUTERS

What we'd like:

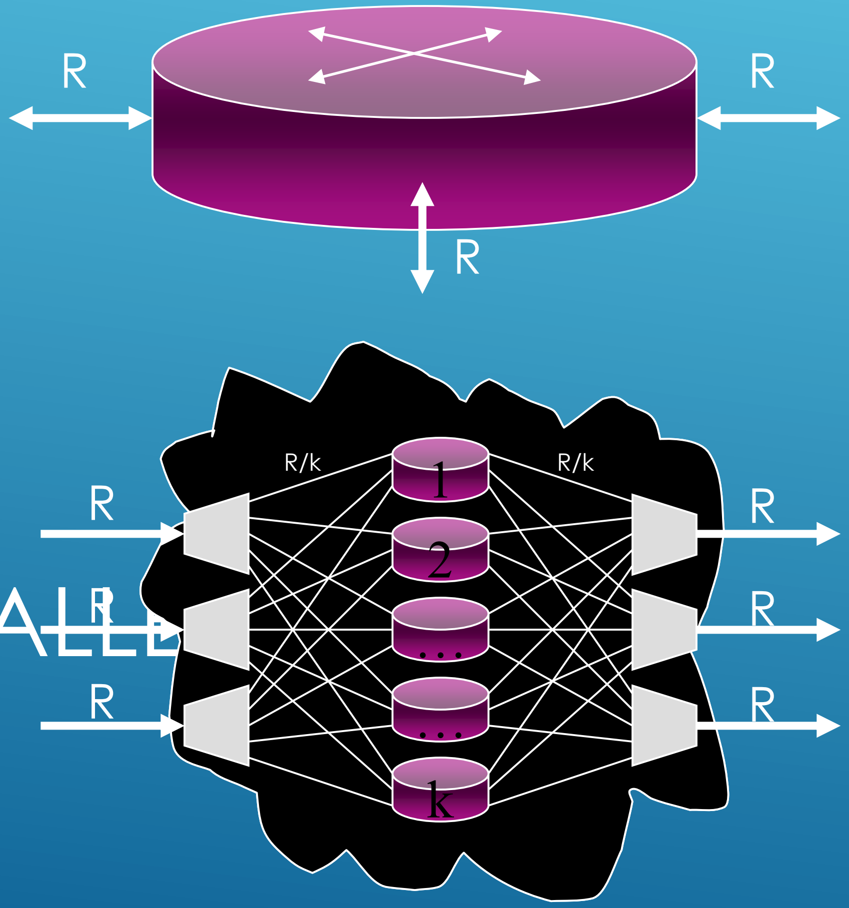


The building blocks we'd like to use:

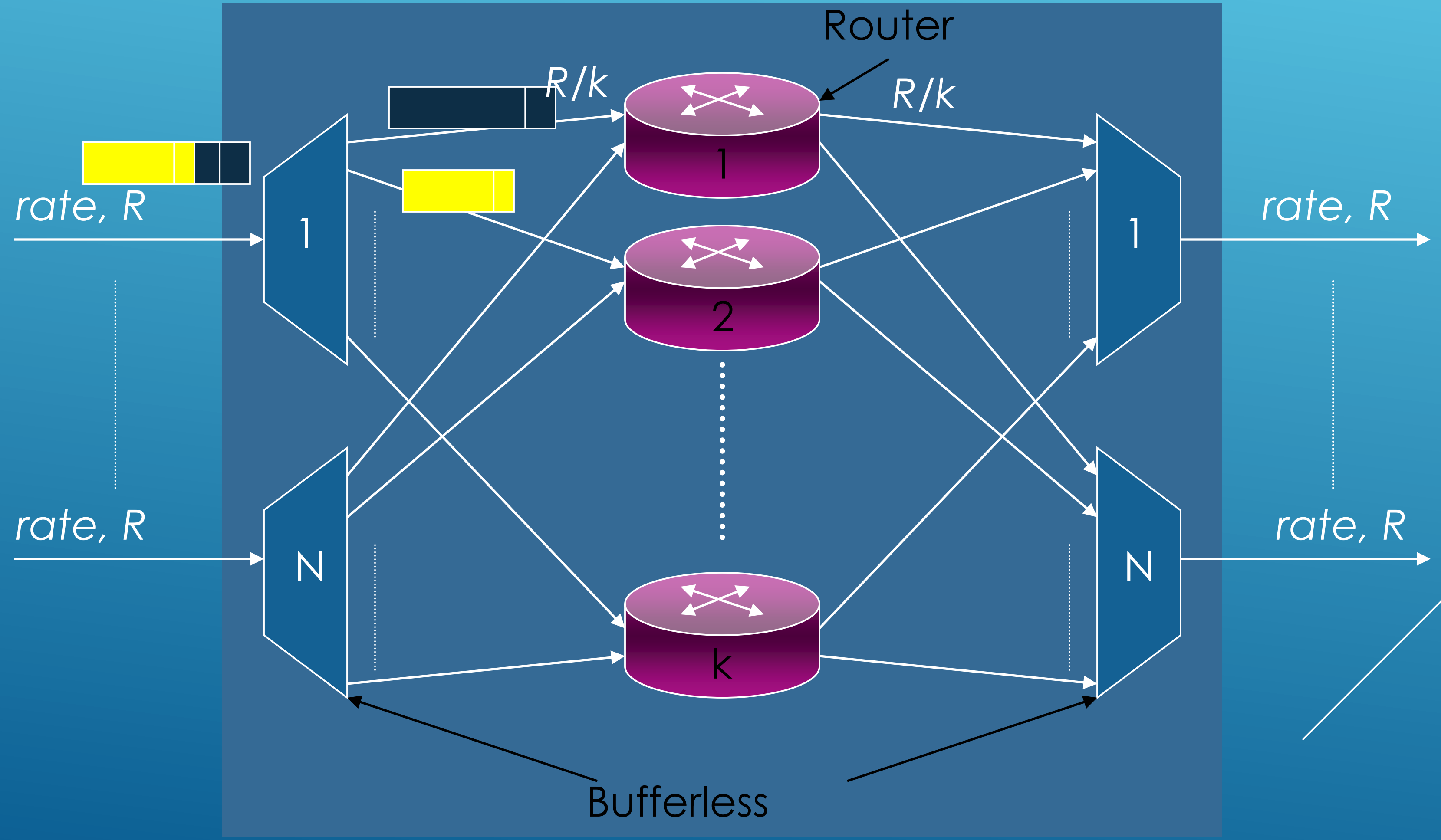


MULTIPLE PARALLE

LOAD BALANCING



INTELLIGENT PACKET LOAD-BALANCING
PARALLEL PACKET SWITCHING



OUTLINE

Background

- What is a router?
- Why do we need faster routers?
- Why are they hard to build?

Architectures and techniques

- The evolution of router architecture.
- IP address lookup.
- Packet buffering.
- Switching.

The Future

- More parallelism.
- Eliminating schedulers.
- Introducing optics into routers.
- Natural evolution to circuit switching?

They are already there.

- Connecting linecards to switches.

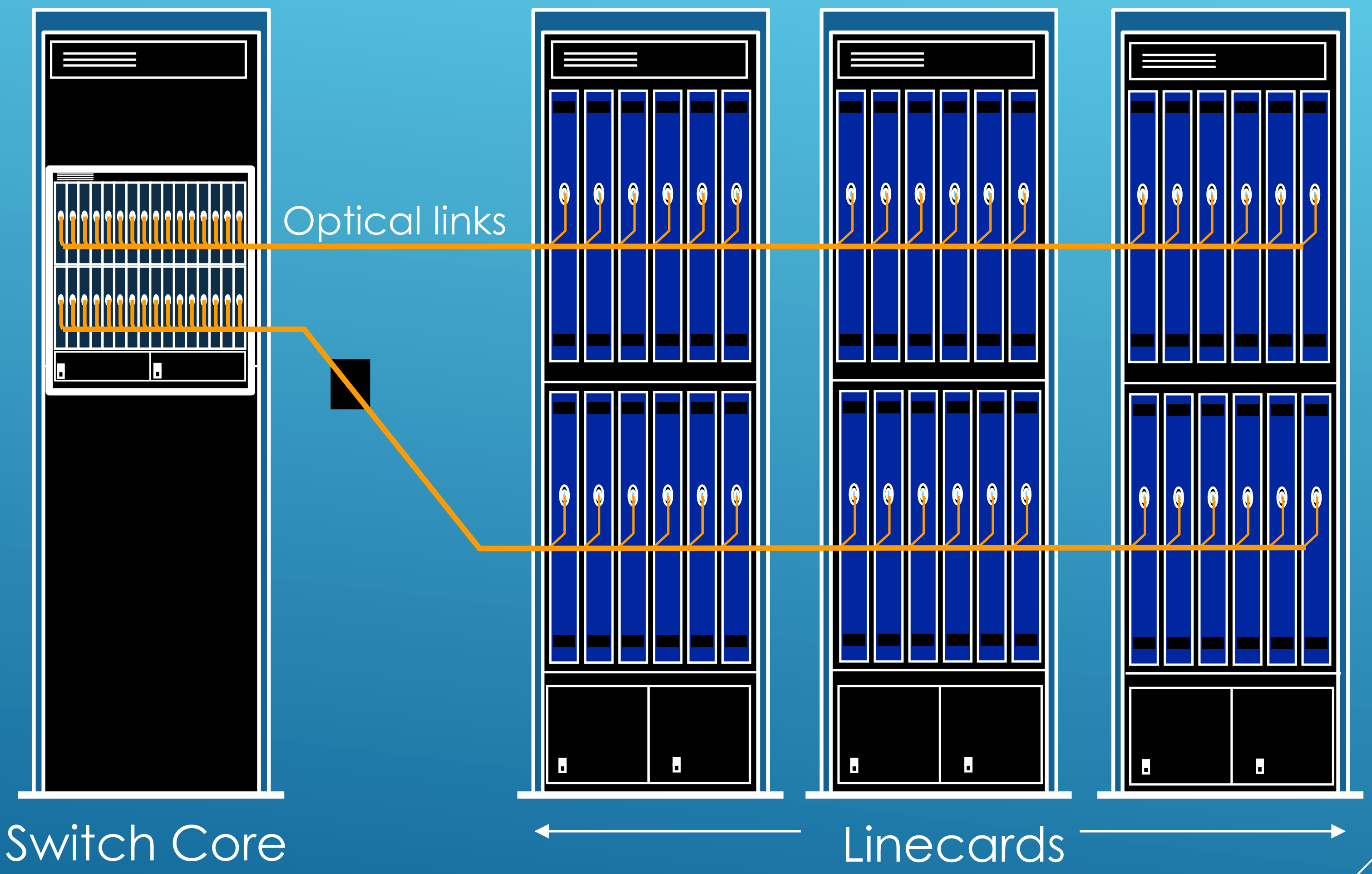
Optical processing doesn't belong on the linecard.

- You can't buffer light.
- Minimal processing capability.

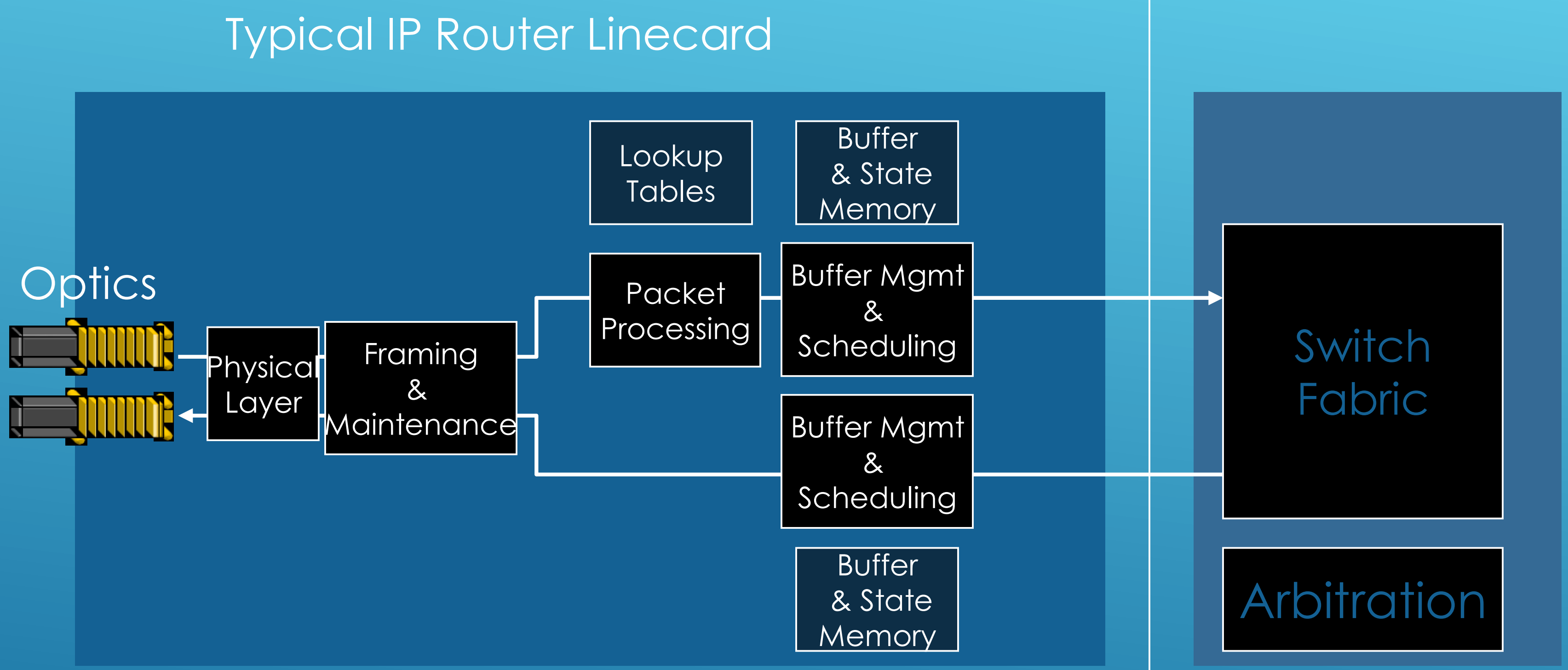
Optical switching can reduce power.

DO OPTICS BELONG IN ROUTERS?

Optics in routers



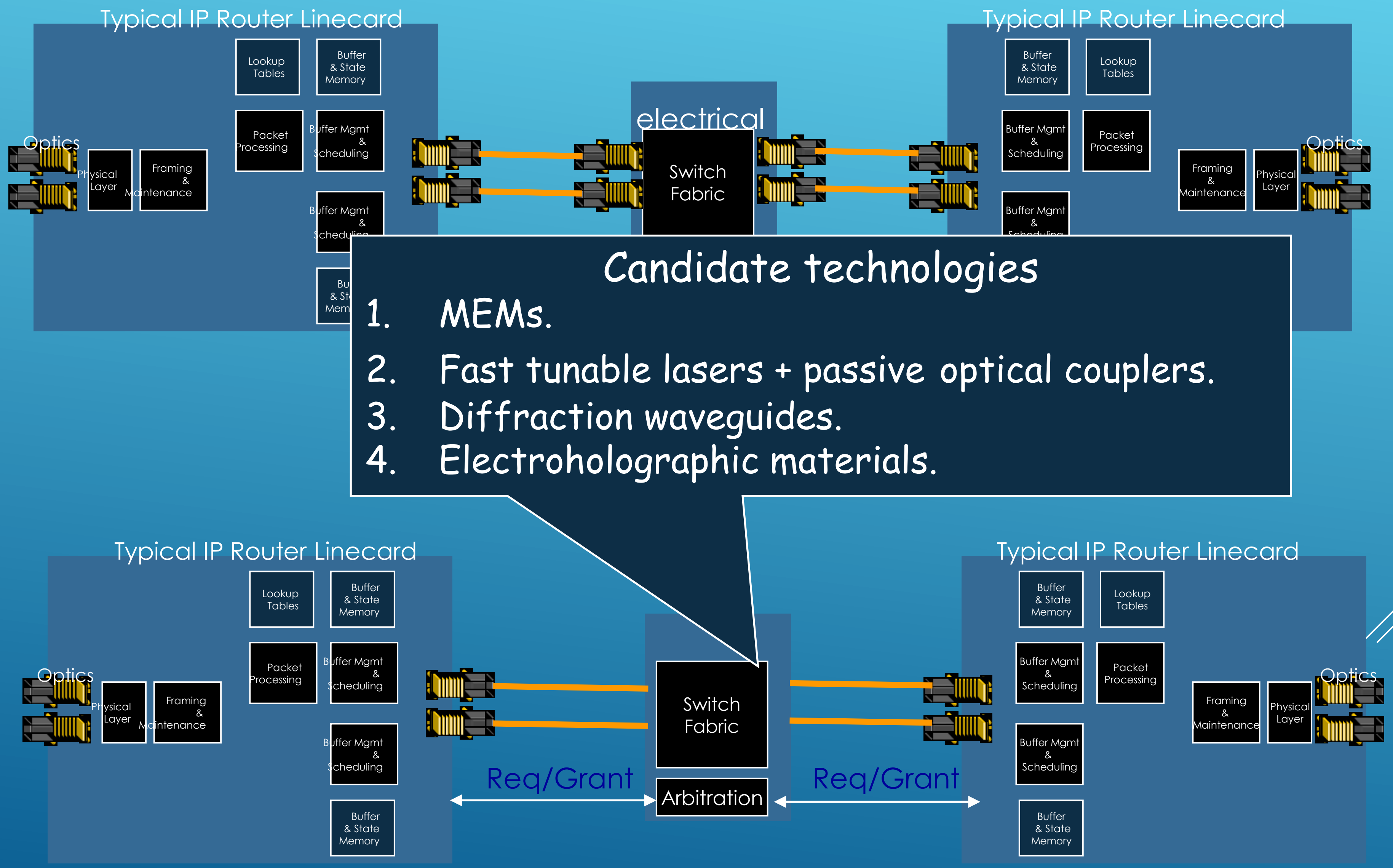
COMPLEX LINECARDS



10Gb/s linecard:

- Number of gates: 30M
- Amount of memory: 2Gbits
- Cost: >$20k
- Power: 300W

REPLACING THE SWITCH FABRIC WITH OPTICS



Candidate technologies

1. MEMs.
2. Fast tunable lasers + passive optical couplers.
3. Diffraction waveguides.
4. Electroholographic materials.

OUTLINE

Background
- What is a router?
- Why do we need faster routers?
- Why are they hard to build?

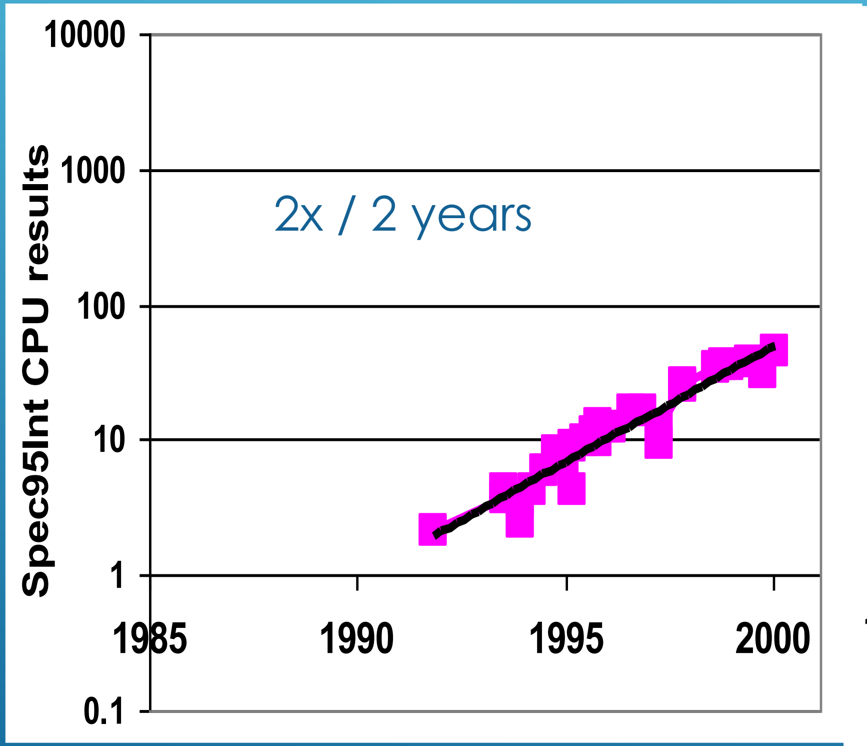
Architectures and techniques
- The evolution of router architecture.
- IP address lookup.
- Packet buffering.
- Switching.

The Future
- More parallelism.
- Eliminating schedulers.
- Introducing optics into routers.
- Natural evolution to circuit switching?

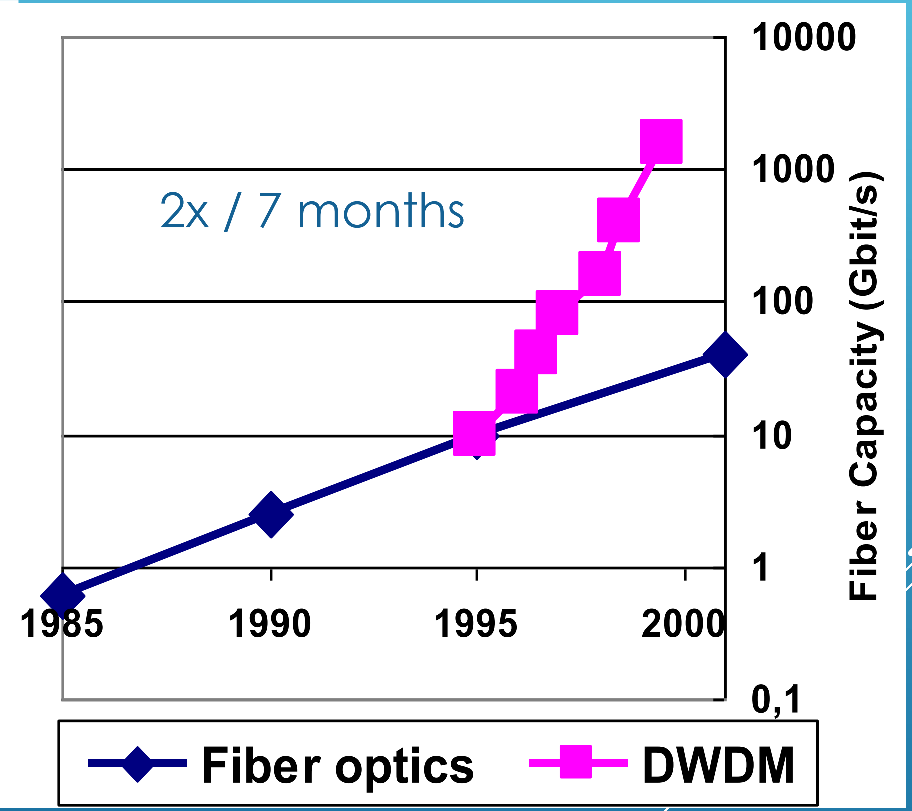
- Optics enables simple, low-power, very high capacity circuit switches.
- The Internet was packet switched for two reasons:
  - Expensive links: statistical multiplexing.
  - Resilience: soft-state routing.
- Neither reason holds today.

EVOLUTION TO CIRCUIT SWITCHING

Processing Power



Link Speed (Fiber)



FAST L

Source: SPEC95Int; Prof. Miller, Stanford Univ.

OUTLINE

Background

- What is a router?
- Why do we need faster routers?
- Why are they hard to build?

Architectures and techniques

- The evolution of router architecture.
- IP address lookup.
- Packet buffering.
- Switching.

The Future

- More parallelism.
- Eliminating schedulers.
- Introducing optics into routers.
- Natural evolution to circuit switching?

REFERENCES

General

1. J. S. Turner "Design of a Broadcast packet switching network", IEEE Trans Comm, June 1988, pp. 734-743.
2. C. Partridge et al. "A Fifty Gigabit per second IP Router", IEEE Trans Networking, 1998.
3. N. McKeown, M. Izzard, A. Mekkittikul, W. Ellersick, M. Horowitz, "The Tiny Tera: A Packet Switch Core", IEEE Micro Magazine, Jan-Feb 1997.

Fast Packet Buffers

1. Sundar Iyer, Ramana Rao, Nick McKeown "Design of a fast packet buffer", IEEE HPSR 2001, Dallas.

REFERENCES

IP Lookups

1. A. Brodnik, S. Carlsson, M. Degermark, S. Pink. "Small Forwarding Tables for Fast Routing Lookups", Sigcomm 1997, pp 3-14.
2. B. Lampson, V. Srinivasan, G. Varghese. " IP lookups using multiway and multicolumn search", Infocom 1998, pp 1248-56, vol. 3.
3. M. Waldvogel, G. Varghese, J. Turner, B. Plattner. "Scalable high speed IP routing lookups", Sigcomm 1997, pp 25-36.
4. P. Gupta, S. Lin, N. McKeown. "Routing lookups in hardware at memory access speeds", Infocom 1998, pp 1241-1248, vol. 3.
5. S. Nilsson, G. Karlsson. "Fast address lookup for Internet routers", IFIP Intl Conf on Broadband Communications, Stuttgart, Germany, April 1-3, 1998.
6. V. Srinivasan, G.Varghese. "Fast IP lookups using controlled prefix expansion", Sigmetrics, June 1998.

REFERENCES

Switching

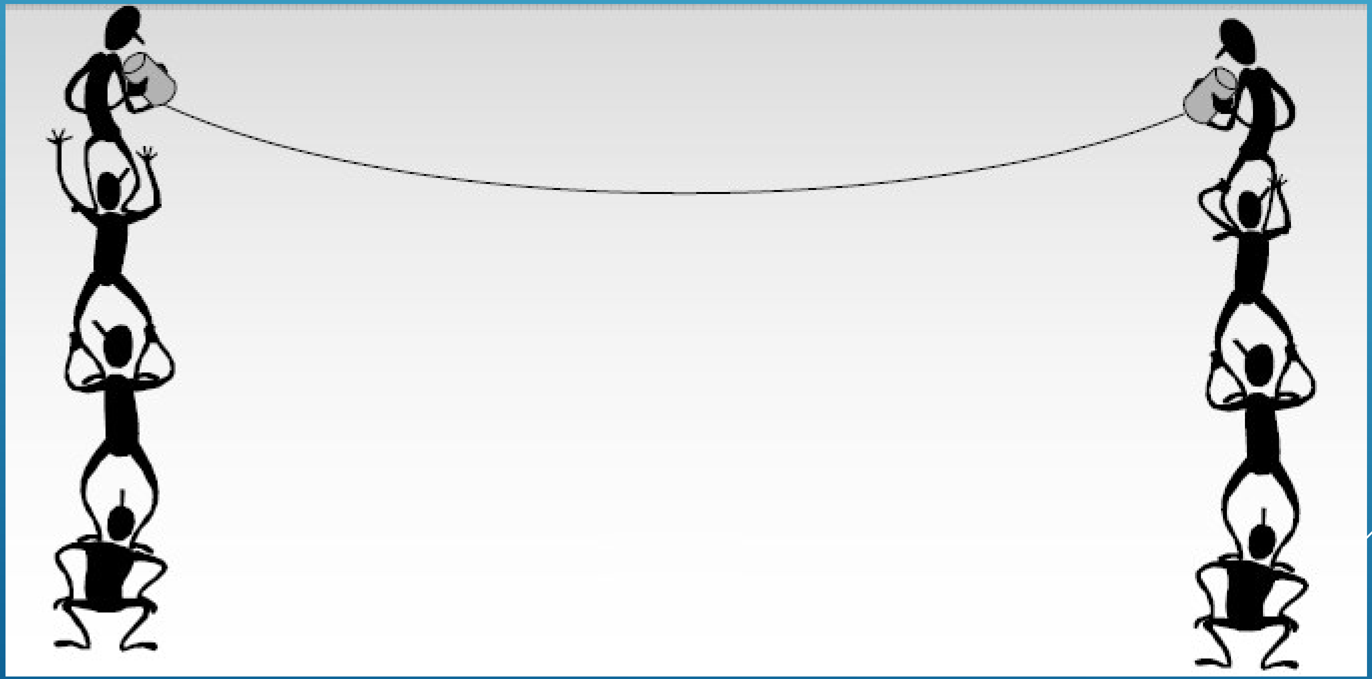
- N. McKeown, A. Mekkittikul, V. Anantharam, and J. Walrand. Achieving 100% Throughput in an Input-Queued Switch. IEEE Transactions on Communications, 47(8), Aug 1999.
- A. Mekkittikul and N. W. McKeown, "A practical algorithm to achieve 100% throughput in input-queued switches," in Proceedings of IEEE INFOCOM '98, March 1998.
- L. Tassiulas, "Linear complexity algorithms for maximum throughput in radio networks and input queued switchs," in Proc. IEEE INFOCOM '98, San Francisco CA, April 1998.
- D. Shah, P. Giaccone and B. Prabhakar, "An efficient randomized algorithm for input-queued switch scheduling," in Proc. Hot Interconnects 2001.
- J. Dai and B. Prabhakar, "The throughput of data switches with and without speedup," in Proceedings of IEEE INFOCOM '00, Tel Aviv, Israel, March 2000, pp. 556 -- 564.
- C.-S. Chang, D.-S. Lee, Y.-S. Jou, "Load balanced Birkhoff-von Neumann switches," Proceedings of IEEE HPSR '01, May 2001, Dallas, Texas.

REFERENCES

Future

- C.-S. Chang, D.-S. Lee, Y.-S. Jou, “Load balanced Birkhoff-von Neumann switches,” Proceedings of IEEE HPSR ‘01, May 2001, Dallas, Texas.
- Pablo Molinero-Fernndez, Nick McKeown "TCP Switching: Exposing circuits to IP" Hot Interconnects IX, Stanford University, August 2001
- S. Iyer, N. McKeown, "Making parallel packet switches practical," in Proc. IEEE INFOCOM `01, April 2001, Alaska.

MULTI PROTOCOL LABEL SWITCHING (MPLS)



WHY INTERNET PROTOCOL IS POPULAR?

- Robustness
- Aggregation and Hierarchy

ISSUES WITH INTERNET PROTOCOL

- IP address lookup
- No QoS
- Best Effort

OBJECTIVES OF MPLS

- Speed up IP packet forwarding
  - By cutting down on the amount of processing at every intermediate router
- Prioritize IP packet forwarding
  - By providing ability to engineer traffic flow and assure differential QoS
- Without losing on the flexibility of IP based network

MPLS – KEY IDEAS

- Use a fixed length label in the packet header to decide packet forwarding
- A path is established between two end points.
- At ingress a packet is classified into a Forwarding Equivalence Class.
- FEC to which it is assigned is encoded as a short fixed length value – Label.
- Packet is forwarded along with label to next hop.

MPLS – KEY IDEAS

- No further analysis of header by subsequent routers, forwarding is driven by the labels.
- Label is used as index for next hop and new label.
- At the Egress, label is removed and packet is forwarded to final destination based on the IP packet header

MPLS HEADER

- Label: 20-bit label value
- Exp: experimental use
  - Can indicate class of service
- S: bottom of stack indicator
  - 1 for the bottom label, 0 otherwise
- TTL: time to live

20	3	1	8
Label	Exp	S	TTL

Forwarding Equivalence Class

- Forwarding Equivalence Class (FEC): A subset of packets that are all treated the same way by an LSR.
- A packet is assigned to an FEC at the ingress of an MPLS domain.
- Subset can be based on
  - Address prefix
  - Host address
  - QoS

Forwarding Equivalence Class

- **Assume packets have the destination address and QoS requirements as**

124.48.45.20	qos = 1	FEC 1	label A
143.67.25.77	qos = 1	FEC 2	label B
143.67.84.22	qos = 3	FEC 3	label C
124.48.66.90	qos = 4	FEC 4	label D
143.67.12.01	qos = 3	FEC 3	label C

LABEL DISTRIBUTION PROTOCOL (LDP)

- **LDP is the set of procedures to inform other LSRs about binding between FEC and label.**
  - **Piggyback on existing protocols(BGP, OSPF,RSVP)**
  - **Separate Label Distribution Protocol**

MPLS LSPS ESTABLISHING

- Static Configuration

  -Operator can provision LSPs by statically configuring label mappings at each LSRs.

MPLS LSPS ESTABLISHING- LDP

- LDP can have two types of neighbors:

1. Directly connected neighbor

   LDP uses UDP hello messages sent on port 646 to all the routers Multicast address (224.0.0.2) to discover directly connected neighbors.

2. Non-directly connected neighbor LDP has the ability to establish LDP session with a router two or more hops away. Hellos in this case are unicast to the peer router using UDP port 646.

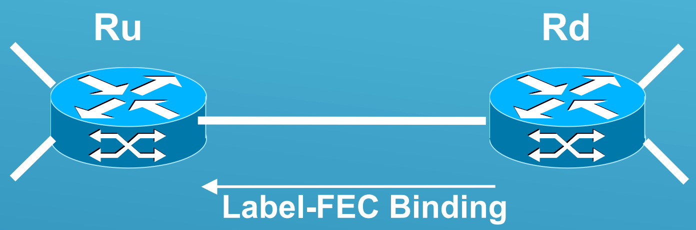
MPLS LSPS ESTABLISHING- LDP

- Once two LSRs discover each other an LDP session is established over which label can be advertised.
- LDP session runs over TCP over port 646.
- Session is maintained by periodic exchange of Keep Alive Messages.

MPLS LSPS ESTABLISHING- RSVP TE

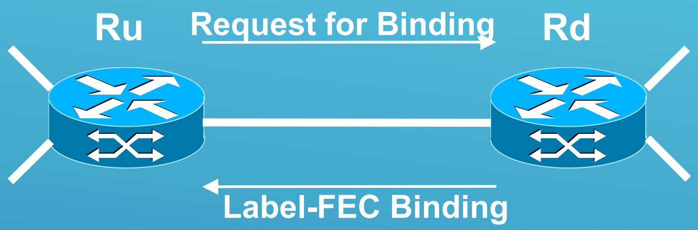
- The ingress LSR computes a path to egress LSR using the CSPF algorithm.
- The computed path is encoded into an Explicit Route Object (ERO) and included in RSVP TE Path messages.
- Each router along the path creates state for the path and forwards the message to the next router in ERO.
- Egress LSR validates the path message and creates a Resv message containing a Label for the LSP.
- The Resv message is sent back to ingress LSR on the same path followed by path message.
- When Resv message reaches the ingress LSR , LSP setup is created

LABEL DISTRIBUTION METHODS



Unsolicited Downstream Label Distribution

- Rd discovers a 'next hop' for a particular FEC
- Rd generates a label for the FEC and communicates the binding to Ru
- Ru inserts the binding into its forwarding tables



Downstream on Demand Label Distribution

- **Ru recognizes Rd as its next-hop for an FEC**
- **A request is made to Rd for a binding between the FEC and a label**
- **If Rd recognizes the FEC and has a next hop for it, it creates a binding and replies to the Ru**

LABEL DISTRIBUTION AND MANAGEMENT

- Label Distribution Control Mode
  - Independent LSP control
  - Ordered LSP control

LABEL DISTRIBUTION AND MANAGEMENT

- Label Retention Mode
  - Conservative – LSR maintains only valid bindings.
  - Liberal - LSR maintains bindings other than the valid next hop, more label, quick adaptation for routing change

LABEL INFORMATION BASE (LIB)

- Table maintained by the LSRs
- Contents of the table
  - Incoming label
  - Outgoing label
  - Outgoing path
- Address prefix

Incoming label	Address Prefix	Outgoing Path	Outgoing label

NEXT HOP LABEL FORWARDING ENTRY (NHLFE)

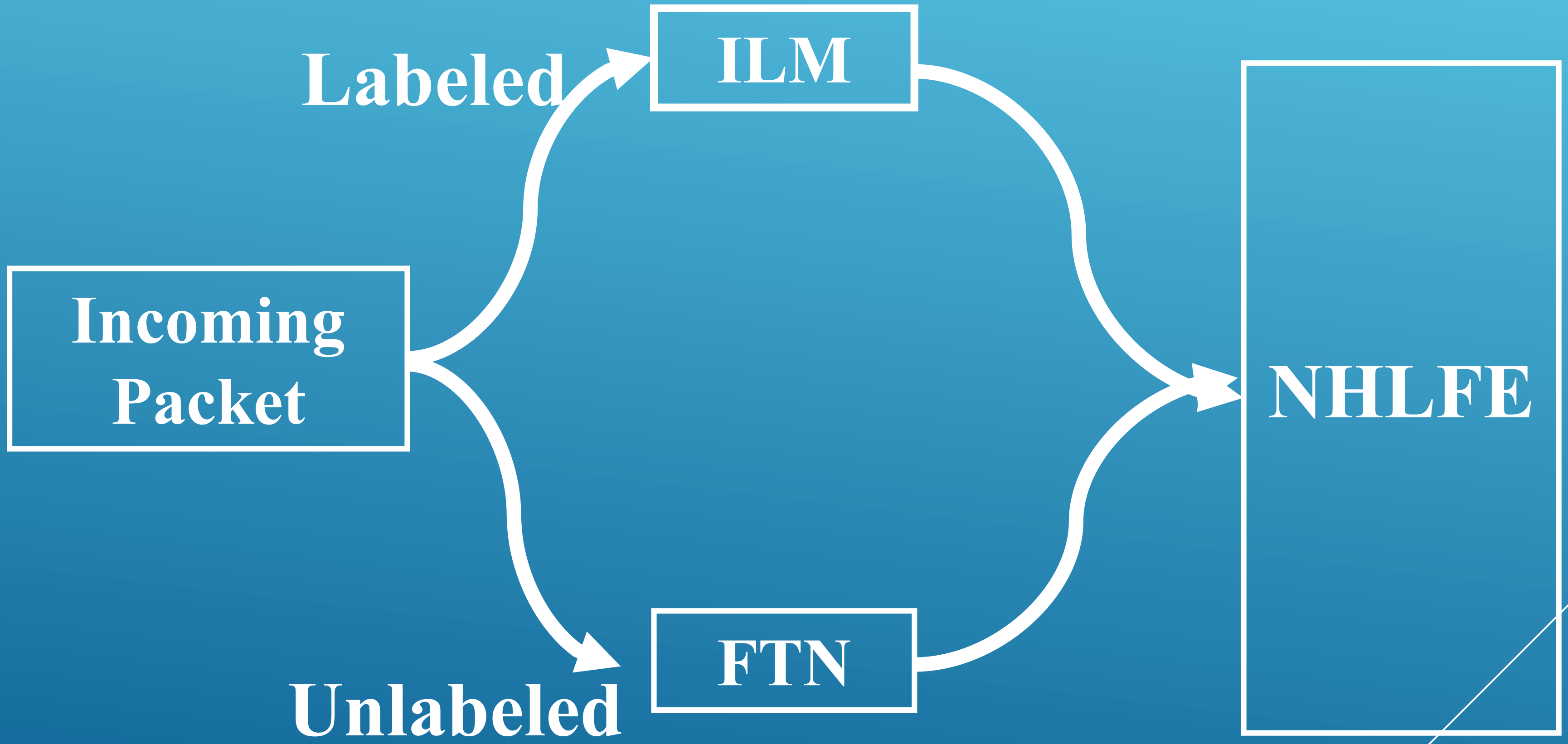
- Used for forwarding a labeled packet.

- Contains the following information:
  - Packets next hop.
  - Operation to be performed on label stack i.e. (One of the following)

➢ **Replace the label on the label stack with a specified new label.**
➢ **Pop the label stack.**
➢ **Replace the label on the label stack with a specified new label and then push one on more specified new label onto the label stack.**

- Incoming Label Map (ILM)- Maps incoming label to NHLFE.
- FEC to NHLFE Map (FTN)-Maps FEC to a set of NHLFE’s.

NHLFE , ILM AND FTN



LABEL STACK

- A labeled packet can contain more than one label.
- Labels are maintained in FIFO stack.
- Processing always done on the top label.
- MPLS support a hierarchy and notion of LSP Tunnel.

OPERATIONS ON LABEL STACK

- Swap
- Pop
- Push

OPERATION ON LABEL STACK-SWAP

- Labeled Packet
- LSR examines label at the top of the label stack of the incoming packet.
- Uses ILM to map to the appropriate NHLFE.
- Encodes the new label stack into the packet and forwards.

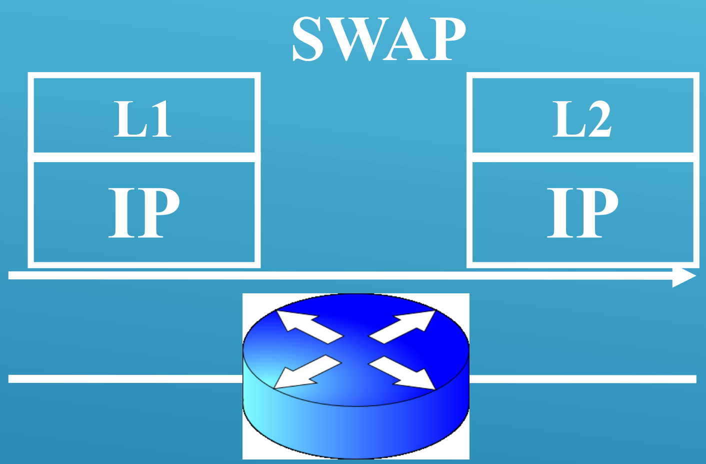
- Unlabeled Packet
- LSR analyses network layer header to determine FEC's.
- Uses FTN to map to an NHLFE.
- Encodes the new label stack into the packet and forwards.

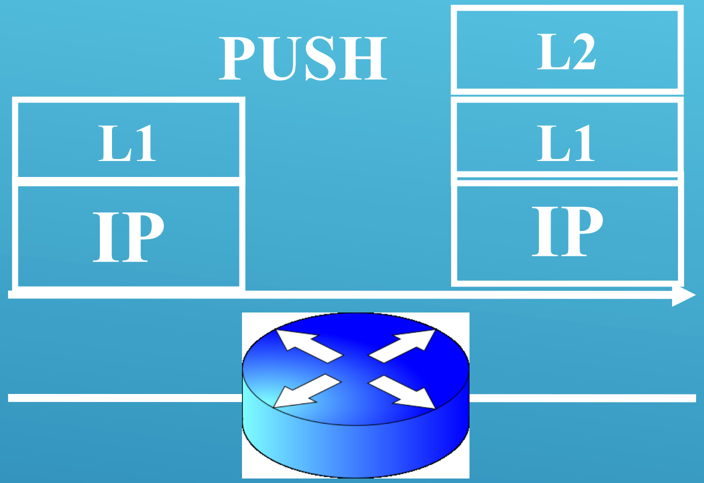
OPERATION ON LABEL STACK

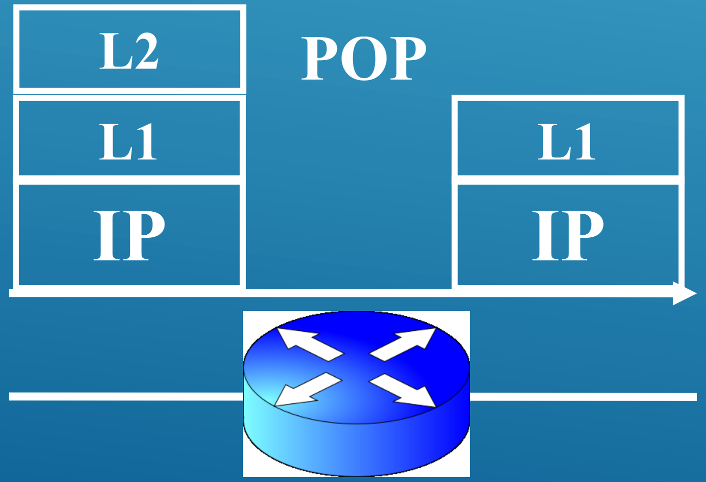
- PUSH
- A new label is pushed on top of the existing label, effectively "encapsulating" the packet in another layer of MPLS.
- This allows hierarchical routing of MPLS packets.

- POP
- The label is removed from the packet, which may reveal an inner label below.
- If the popped label was the last on the label stack, the packet "leaves" the MPLS tunnel.
- This is usually done by the egress router and in Penultimate Hop Popping.

OPERATION ON LABEL STACK







Label Switched Path

- **For each FEC, a specific path called *Label Switched Path (LSP)* is assigned.**
- **To set up an LSP, each LSR must**
  - **Assign an incoming label to the LSP for the corresponding FEC**
  - **Inform the upstream node of the assigned label**
  - **Learn the label that the downstream node has assigned to the LSP**
- **Need a label distribution protocol so that an LSR can inform others of the label/FEC bindings it has made.**
- **A forwarding table is constructed as the result of label distribution.**

Label Switched Path

- **Penultimate Hop Popping : Label Stack is popped at the penultimate LSR of LSP rather than LSP egress.**

✓ **Egress LSP does a single look up.**

✓ **Egress may not be a a LSR.**

Label Switched Path

LSP Next Hop

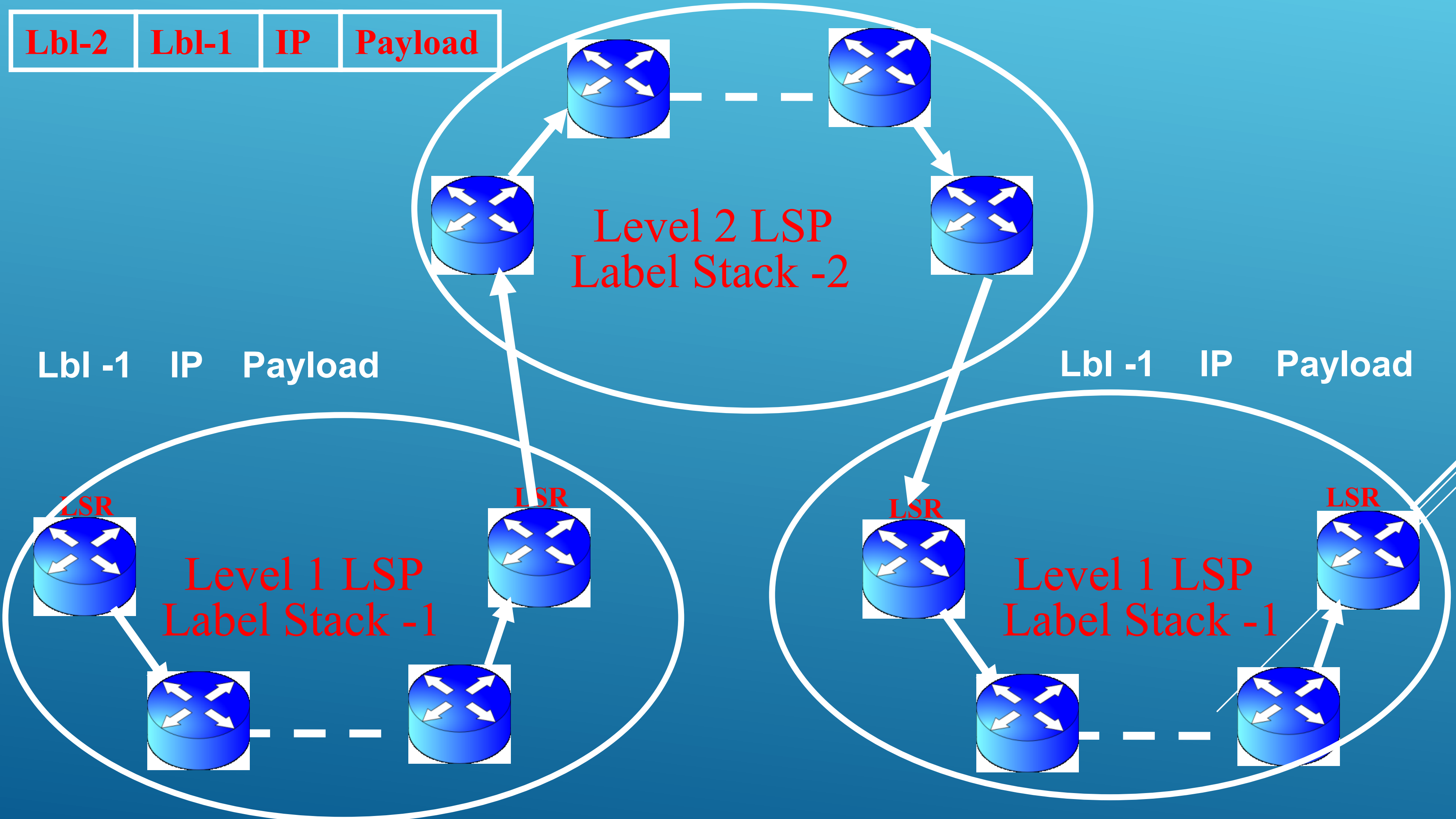
- **Labeled Packet**

- **As selected by the NHLFE entry used for forwarding that packet.**

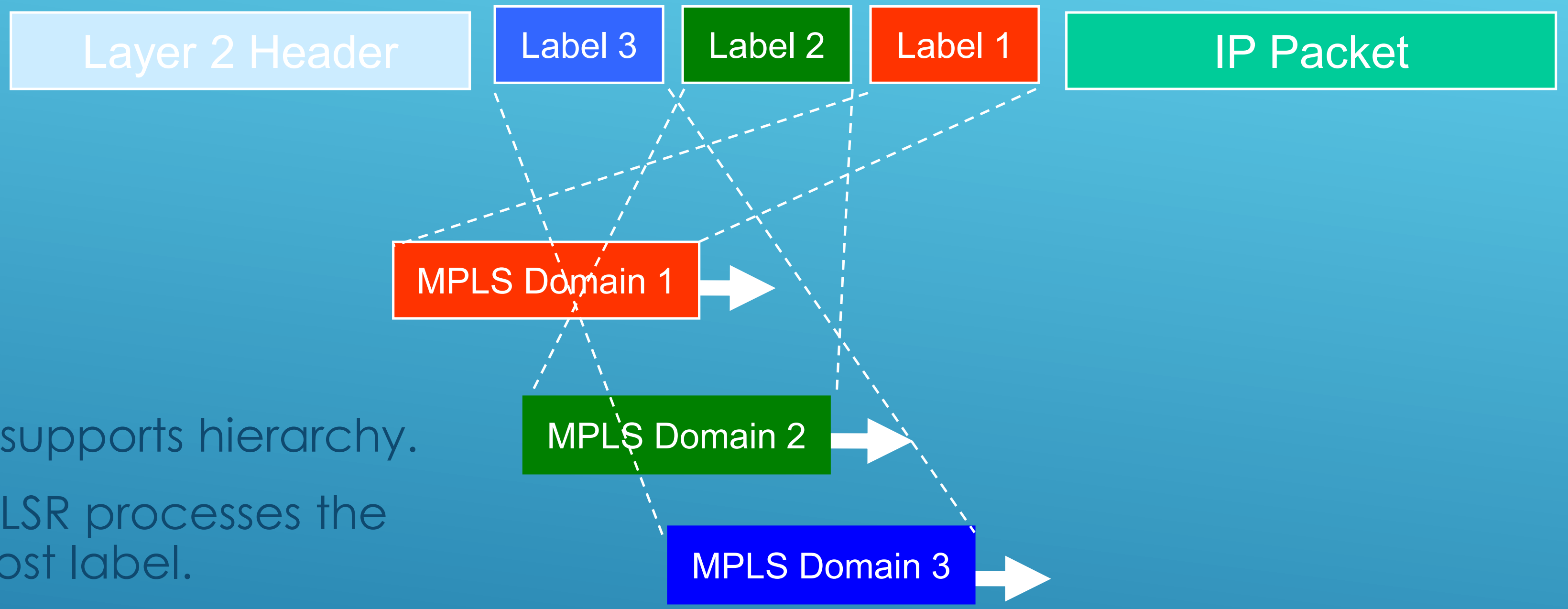
- **FEC**

- **As selected by the NHLFE entry indexed by a label corresponding to that FEC.**

LABEL SWITCHED PATH -HIERARCHY

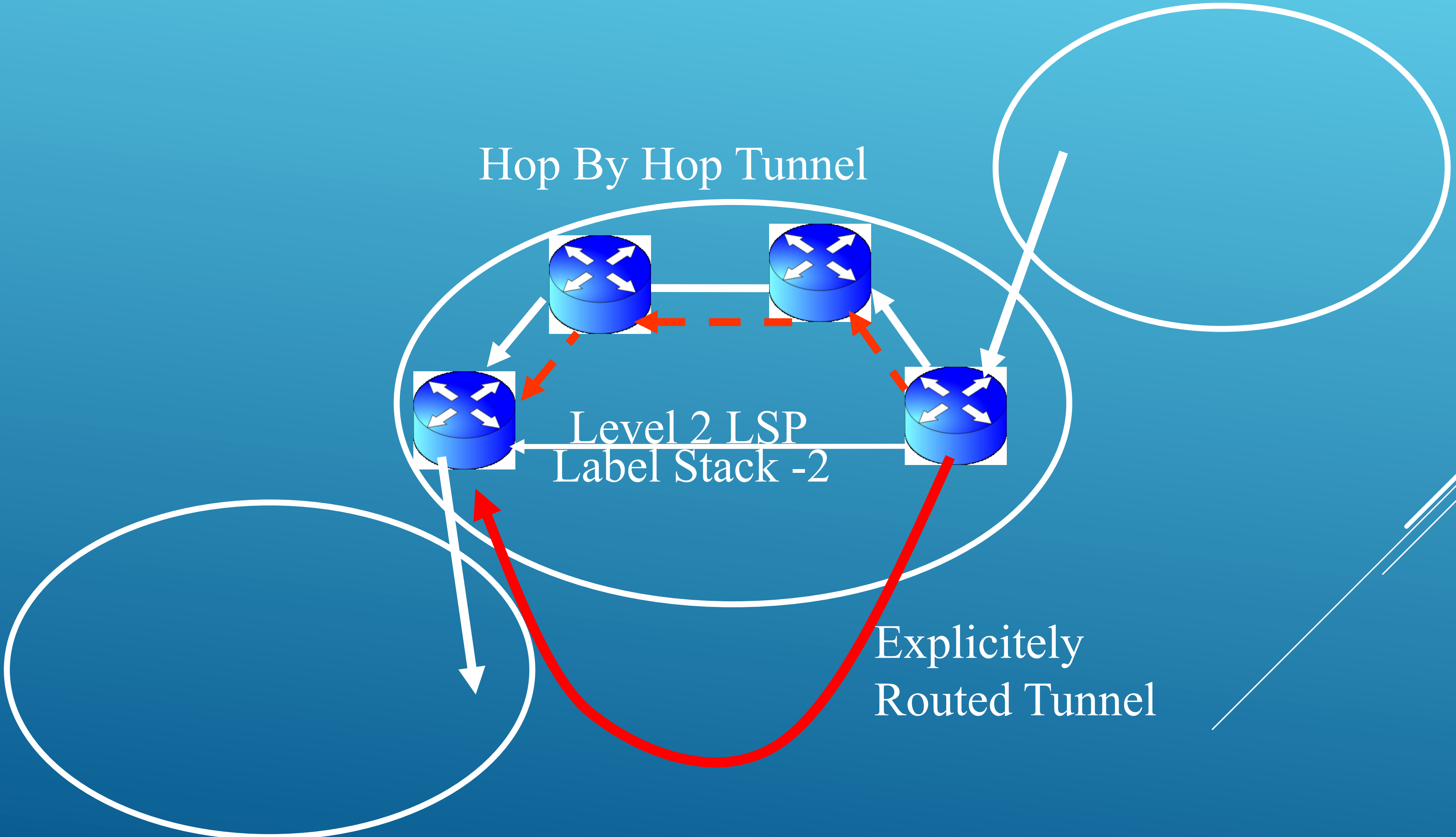


LABEL STACK - HIERARCHY



- MPLS supports hierarchy.
- Each LSR processes the topmost label.
- If traffic crosses several networks, it can be tunneled across them
- Advantage – reduces the LIB table of each router drastically

LABEL SWITCHED PATH -TUNNEL



Label Switched Path - Control

- **Independent**

- **Each LSR on recognizing a particular FEC makes an independent decision to bind a label to it and distribute that binding.**

- **Ordered**

- **An LSR binds a label to a FEC only if it is the egress LSR to that FEC or it has already a binding for that FEC from its next hop for that FEC.**

LSP Route Selection

- **Method for selecting the LSP for a particular FEC.**

- **Hop-by-hop routing: Each node independently choose the next hop for a FEC.**

- **Explicit routing (ER): the sender LSR can specify an *explicit route* for the LSP**
  - **Explicit route can be selected ahead of time or dynamically**

Explicitly Routed LSP

- **Advantages**
  - **Can establish LSP's based on policy, QoS, etc.**
  - **Can have pre-established LSP's that can be used in case of failures.**
  - **It makes MPLS explicit routing much more efficient than the alternative of IP source routing.**

AGGREGATION

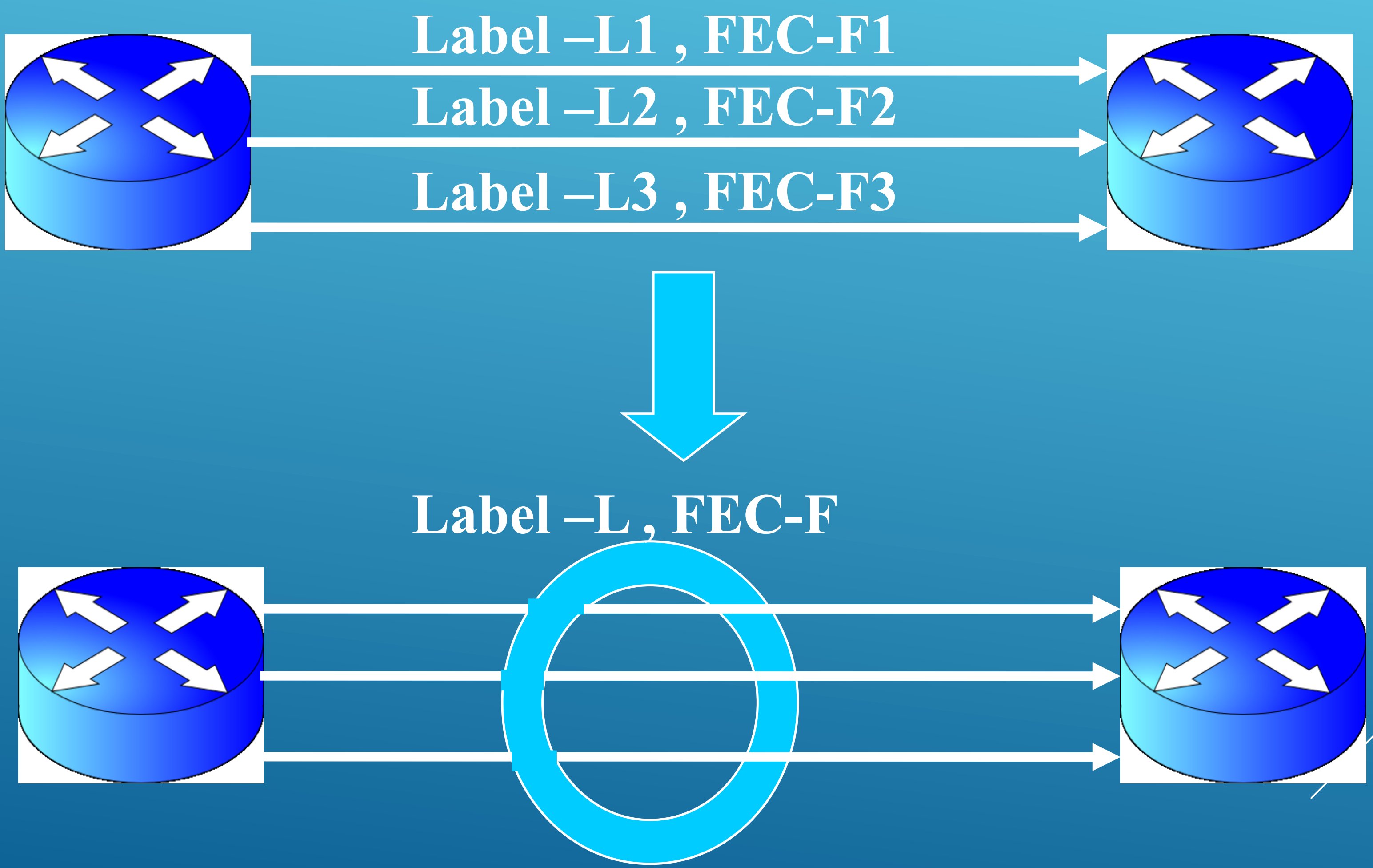
- In the MPLS Domain, all the traffic in a set of FECs might follow the same route.

- The procedure of binding a single label to a union of FECs which is itself a FEC, and applying that label to all traffic in the union is known as Aggregation.

✓ Reduces the number of labels.

✓ Reduces the amount of label distribution control traffic.

AGGREGATION



LABEL MERGING

- *Label Merging* is the capability of forwarding two different packets belonging to the same *FEC*, but arriving with different *labels*, with the same *outgoing label*.

- An *LSR* is *label merging capable* if it can receive two packets from different incoming interfaces, and/or with different labels, and send both packets out the same outgoing interface with the same label.

- Once transmitted, the information that they arrive from different interfaces and/or with different labels is lost.

MPLS PROTECTION

- MPLS OAM*: Fault detection and diagnosis
- TRAFFIC PROTECTION: Route traffic away from failed node/link.
- NODE PROTECTION: Enhance node availability.

* Operation, Administration and Maintenance.

SOME MPLS TRANSPORT PROBLEMS

- Data plane fails ("Black Holes").
- Connectivity Problem, Broken link.
- What path is being taken?

LEVERAGING MPLS OAM

- Difficult to detect MPLS failure:

    Traditional ping may not be successful

- Difficult to troubleshoot MPLS failure:

    Manual hop/hop work

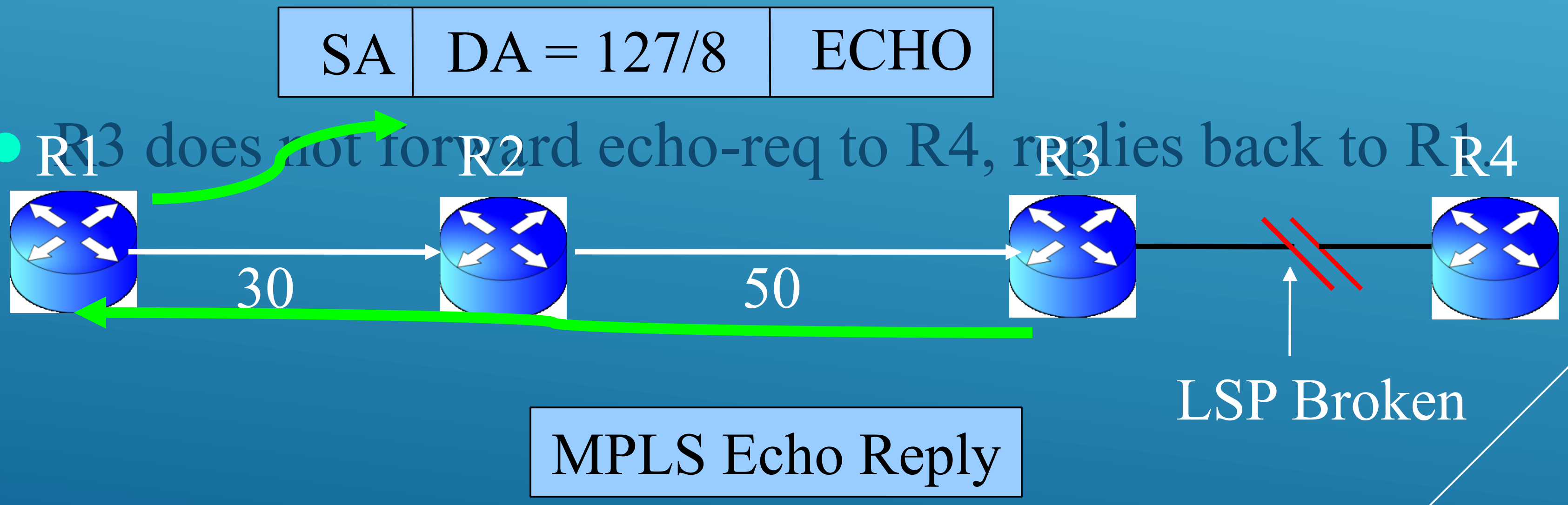
- MPLS OAM facilitates and speeds up troubleshooting of MPLS failures.

LSP PING/TRACEROUTE

- Requirements:
  - Detect MPLS Black holes.
  - Isolate MPLS faults.
  - Diagnose connectivity problems.
- Solutions:
  - LSP ping for connectivity checks.
  - LSP traceroute for hop-by-hop fault localization.
  - LSP traceroute for path tracing.

TROUBLESHOOTING

- R3 does not forward echo-req to R4, replies back to R1



TRAFFIC PROTECTION

- Detect the fault.
- Divert the traffic away from fault.
- Mechanisms:

  - LDP signaled LSPs

  - Backup LSP

  - Fast Reroute

LDP SIGNALED LSPS

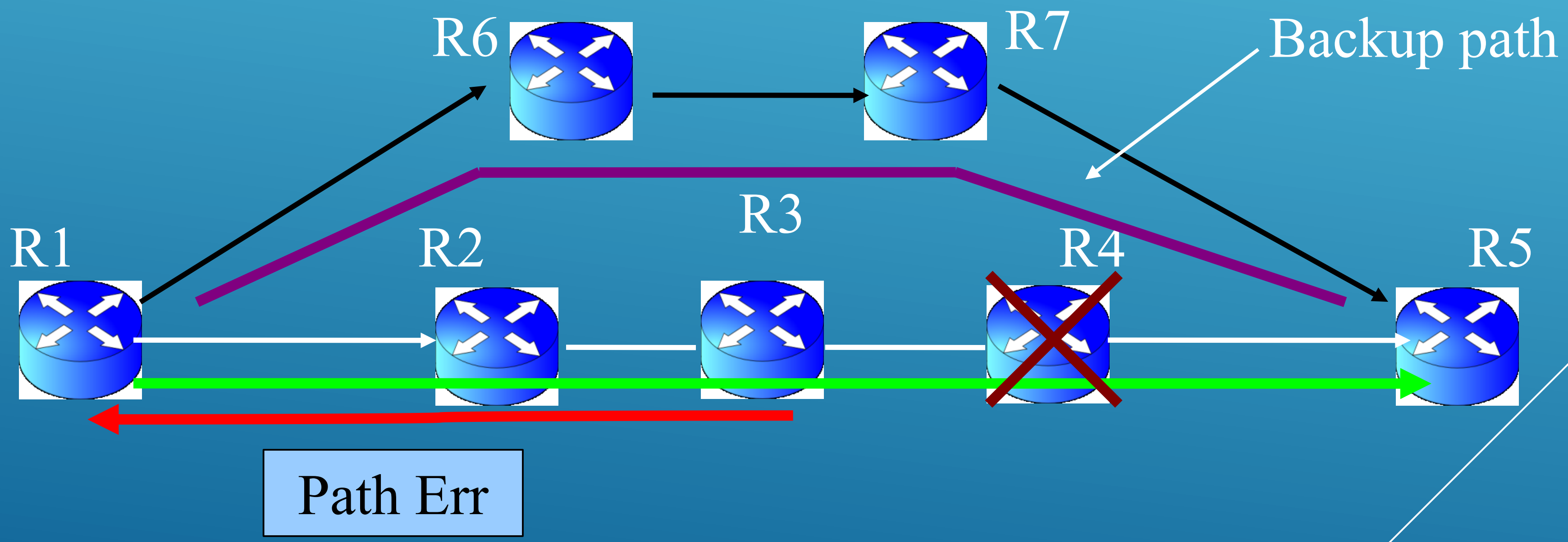
- Path protection depends on IGP reconvergence.
- In case of link/node failure:

i) Remove label mapping for the LSP from FIB.

ii) Wait for new shortest path and matching label.

BACKUP LSPS

- **Backup Path for RSVP-TE signaled LSP**



FAST REROUTE

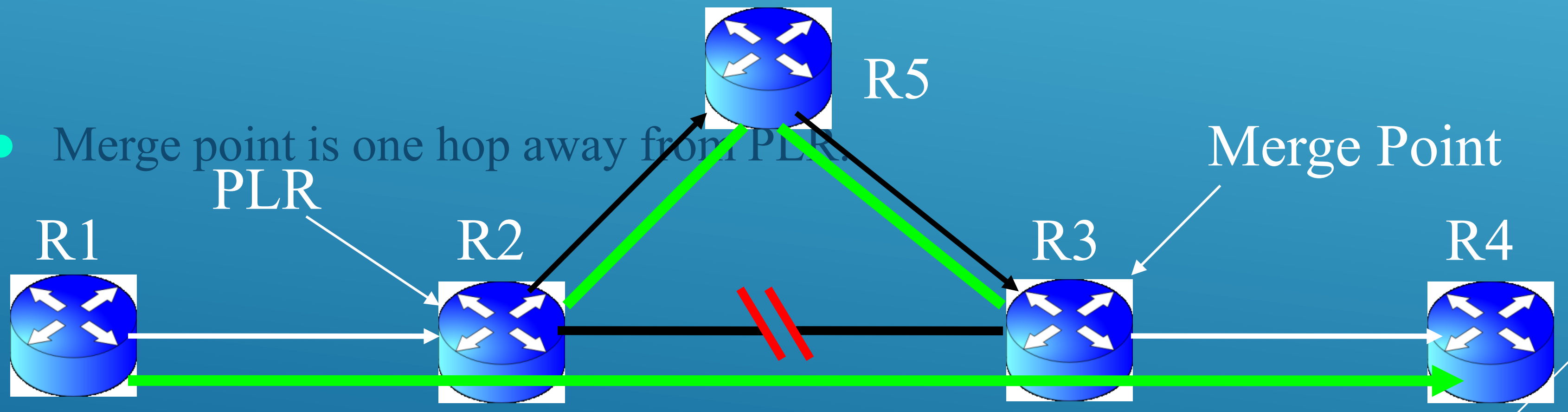
- Repair failure at the node that detects failure (Point of Local Repair or PLR).
- One to one backup:

  Each LSR creates a detour LSP for each protected LSP.
- Facility Bypass:

  Single bypass LSP for all protected LSPs

LINK PROTECTION(FR)

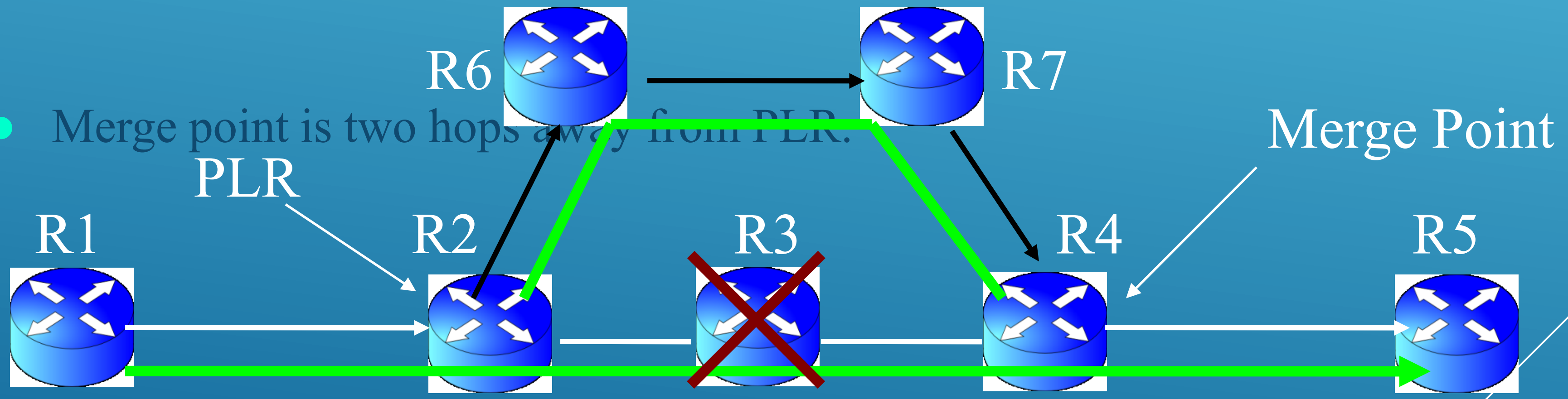
- Merge point is one hop away from PLR.



Next Hop protection

NODE PROTECTION(FR)

- Merge point is two hops away from PLR.



Next-Next Hop protection

NODE PROTECTION

- Resilient LSRs.
- Software support required to take advantage of Hardware redundancy.
- Nonstop Routing:

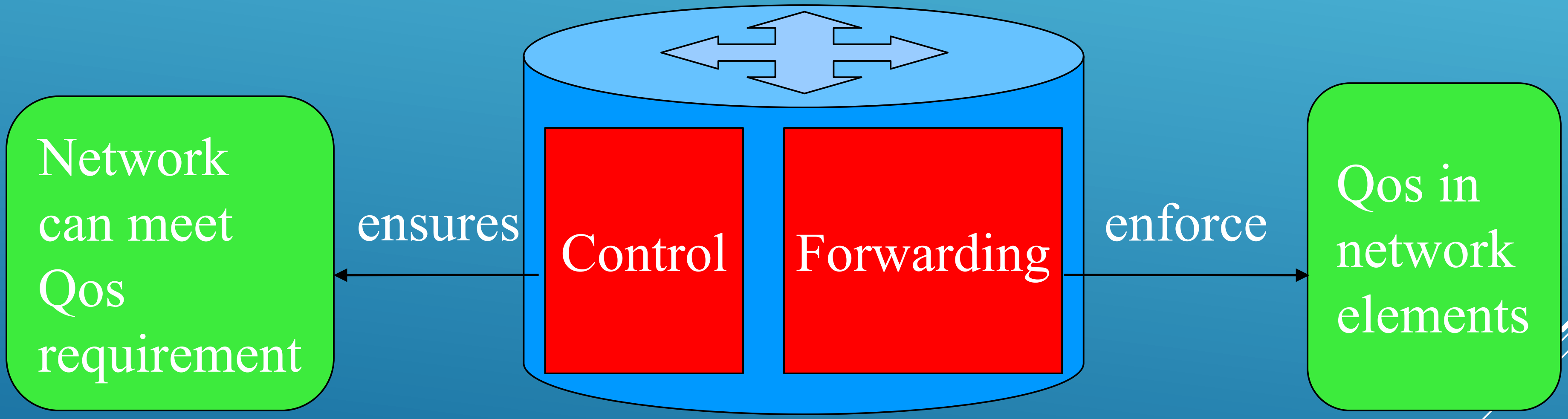
  - Replicate all state changes to backup control card.

  - State synchronization required.
- Nonstop Forwarding:

  - Graceful restart of the failed node.

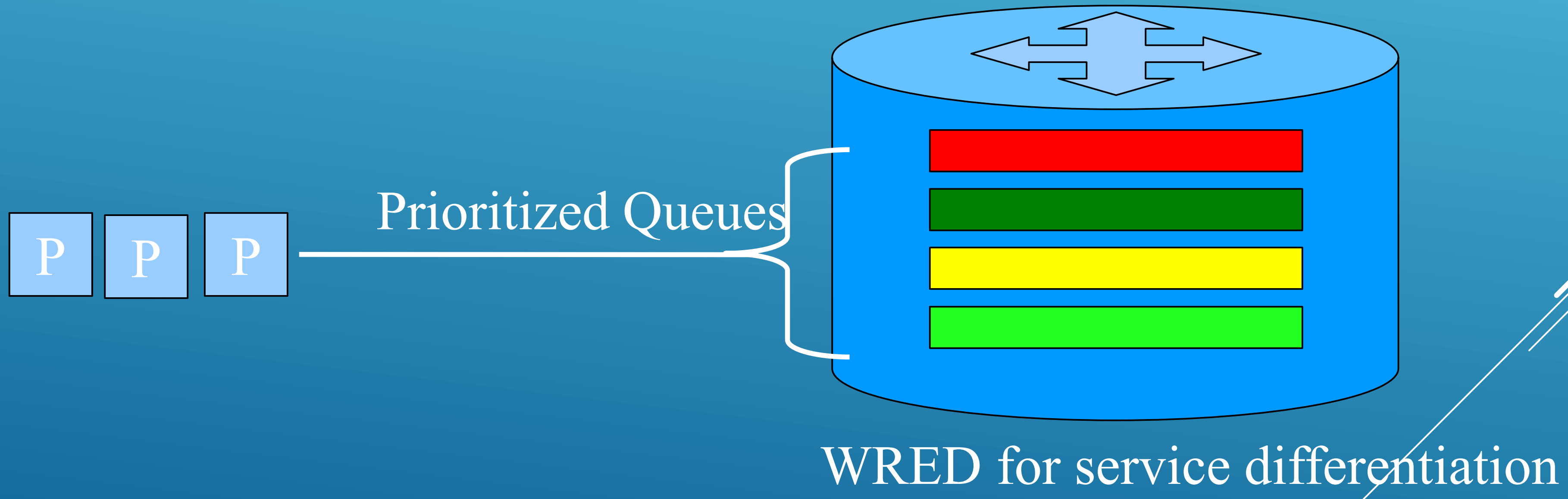
  - Protocol extensions at neighbors.

MPLS QOS



Network can meet Qos requirement
ensures
Control
Forwarding
enforce
Qos in network elements

MPLS QOS



MPLS QOS MODELS

Soft QOS (Class of Service):

- Forwarding plane technique.
- No need for per-flow state (Scales well)
- Unable to provide guaranteed forwarding behavior.

Hard QOS :

- Resources reserved using control plane.
- Per-flow state required.
- Provides firm guarantees.

TRANSPORT PLANE MODELS

- E-LSP model

  - Implementation of Soft QOS model.

  - EXP field in MPLS label.

  - 3 bits => up to 8 different DiffServ points.

- L-LSP model

  - Implementation of Hard QOS model.

  - Both Label and EXP field are considered.

  - Label lookup and Exp bits determine output queue and priority .

TRANSPORT NETWORKS:
THE SERVICE PROVIDER PERSPECTIVE…

Shift from *best effort* to ***guaranteed services***.

Focus on *revenue bearing services*.

Support for varied application requirements.

- Flexible on-demand service granularity.
- High QoS, Reliability.
- Operations, Administration, and Maintenance features.

Low CAPEX and OPEX networks

POSSIBLE SOL

Efficiency and Flexibility

PACKET TECHNOLOGIES
(e.g. IP/MPLS, Carrier Ethernet)

Service Guarantees with Reliability

CIRCUIT TECHNOLOGIES
(e.g. SONET/SDH)

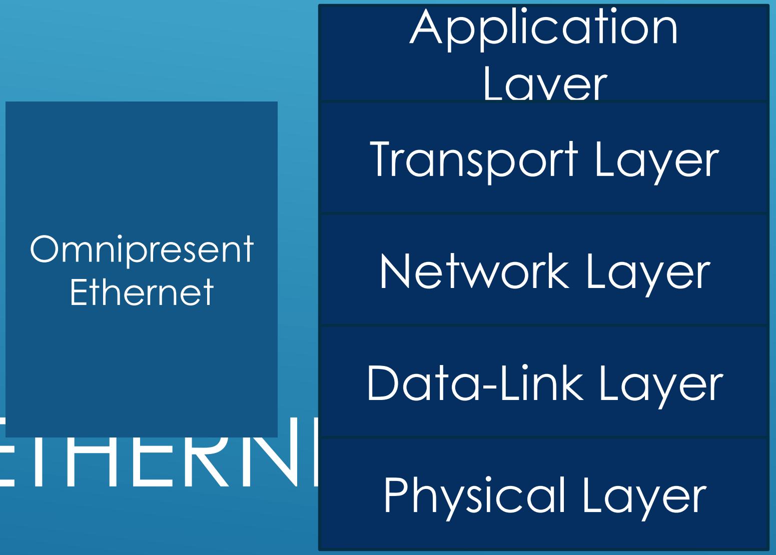
Packet-Optical Transport System

(P-OTS)

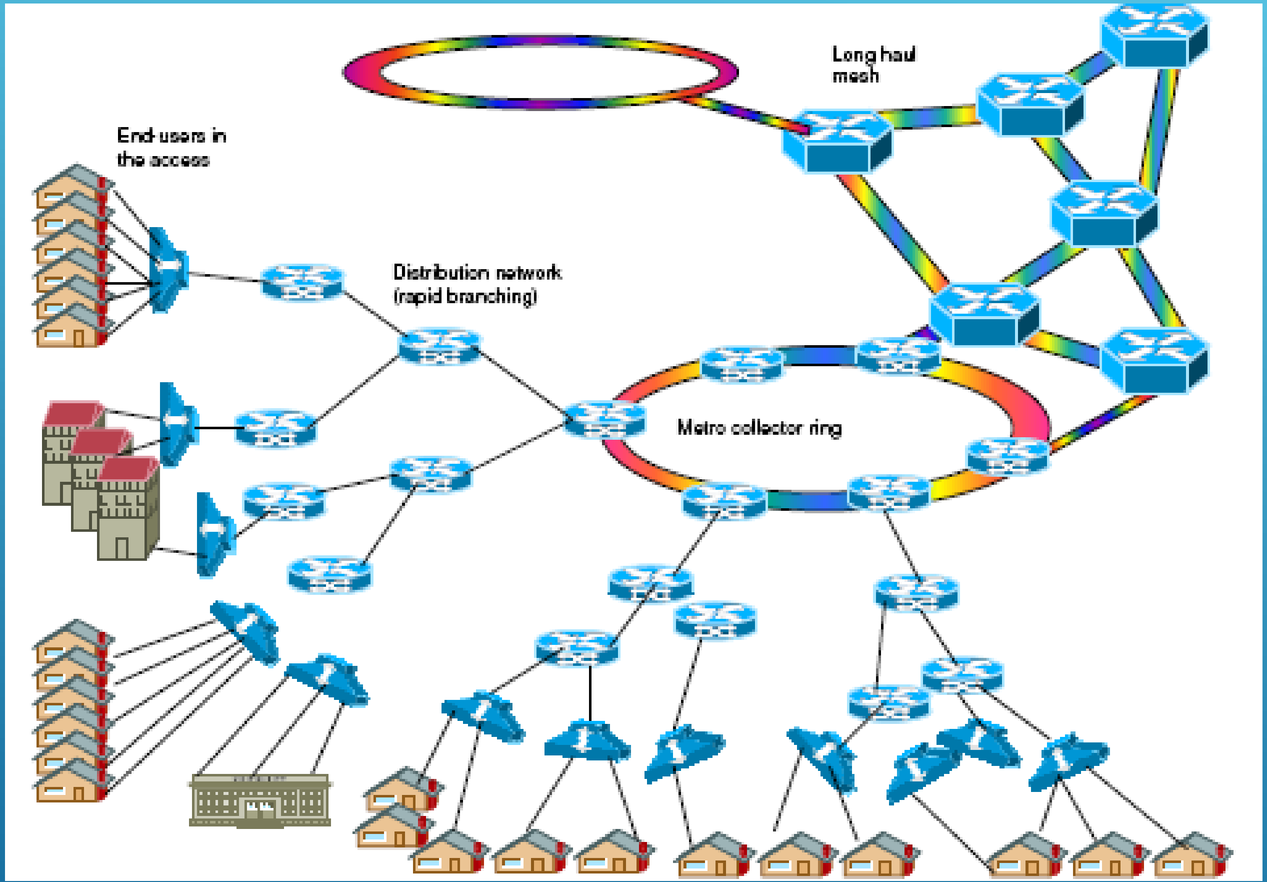
Multi-Layer Optimized transport for **flexible**, **reliable** and **scalable** networks

OMNIPRESENT ETHERNET

- Very fast communication framework combining switching, routing and transport in single layer



CONTEMPORARY NETWORKING HIERARCHY



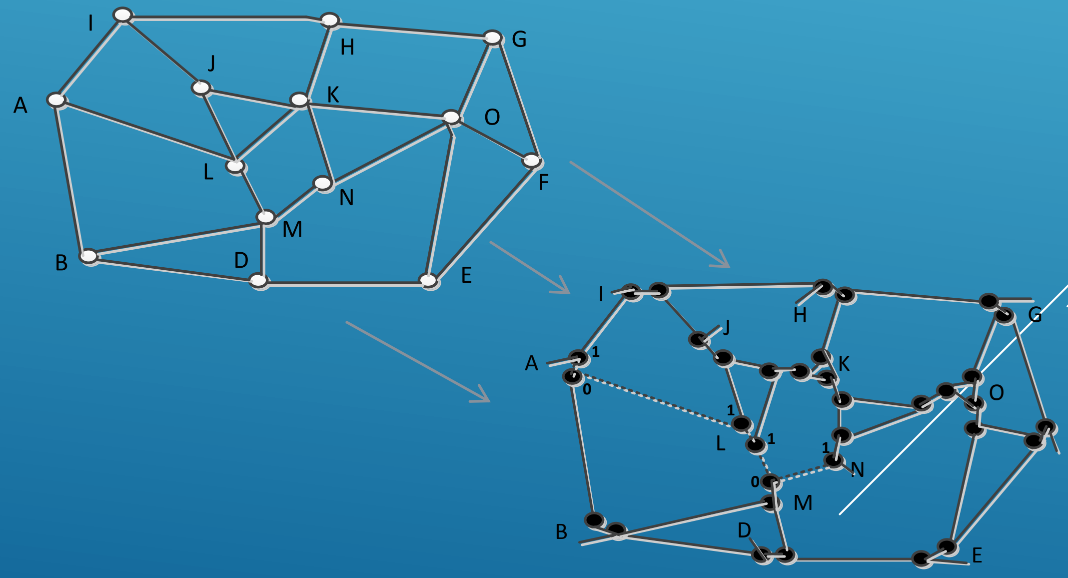
Access, Metro and core networks

- Formulate physical network into a logical hierarchical tree
  - Simplify the network into a fractal binary tree

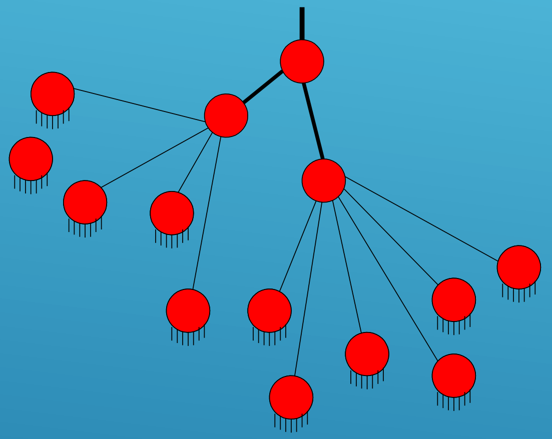
Fig. 1a. Physical topology of the network.

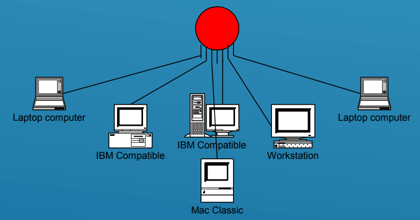
- Binary Routing
  - Packet switching based on the bit value corresponding to the node – *'0'* indicates right-ward movement while a *'1'* indicates left-ward movement

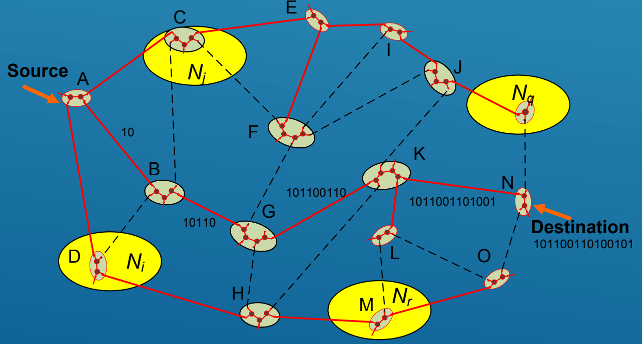
- Advantages:
  - Simple lookup
  - Energy efficient
  - *Source routing*
  - *Cost, cost, cost!!!!!*

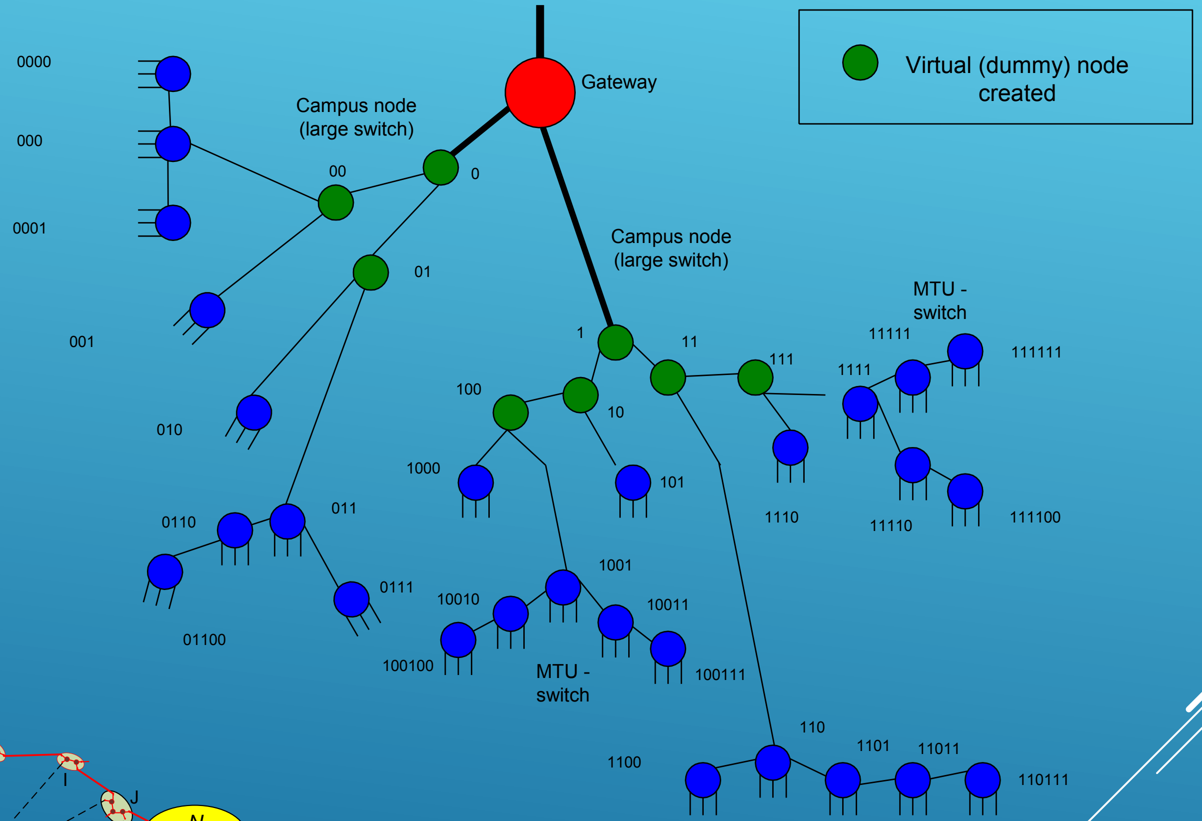


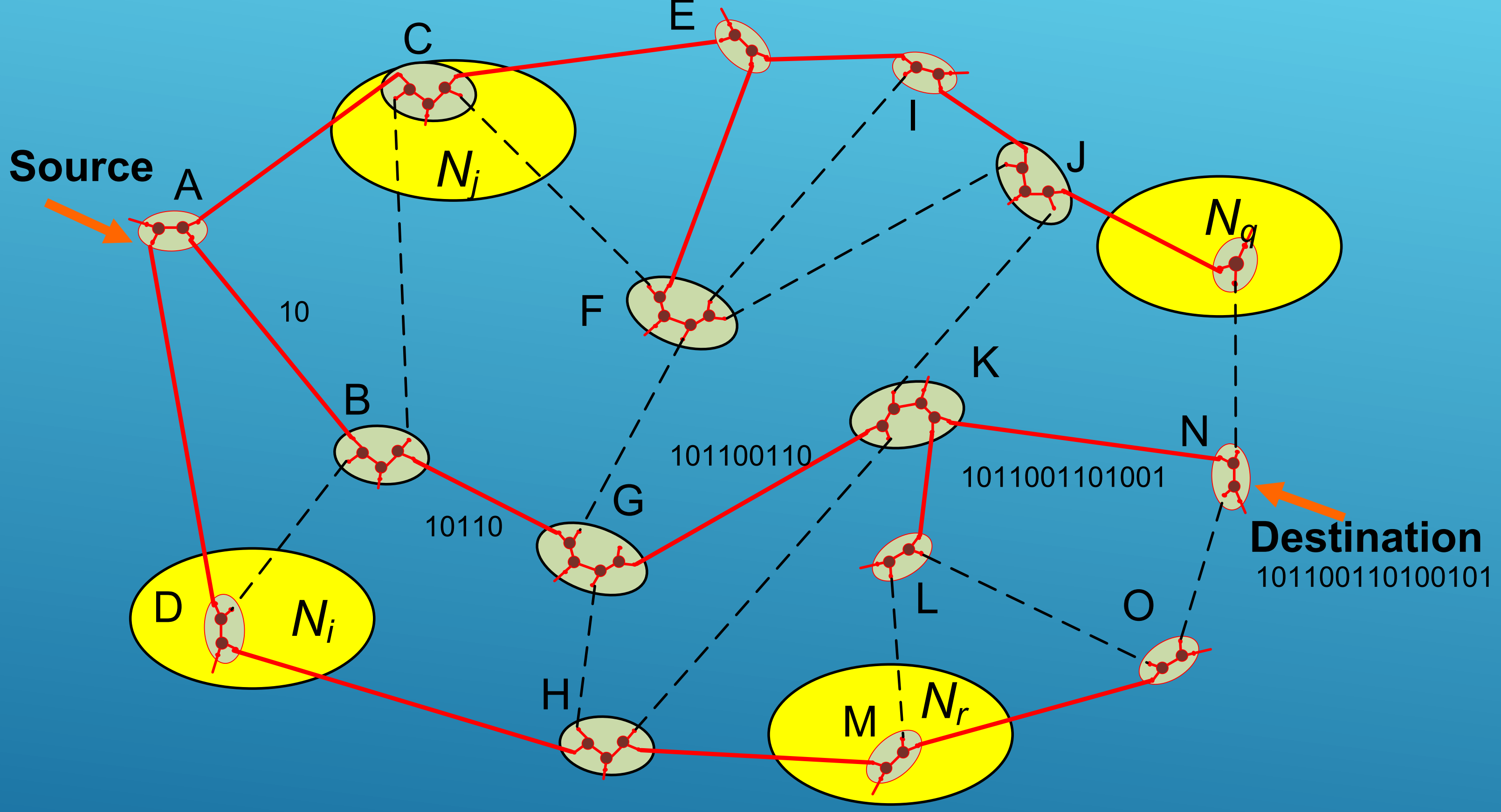
EXAMPLE OF CONVERSION TO A BINARY TREE



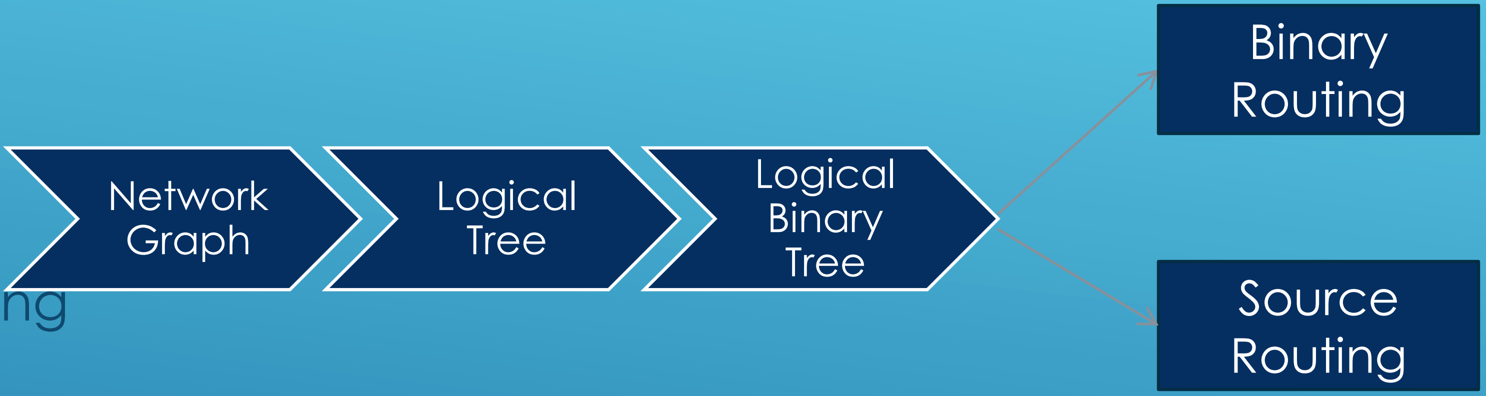






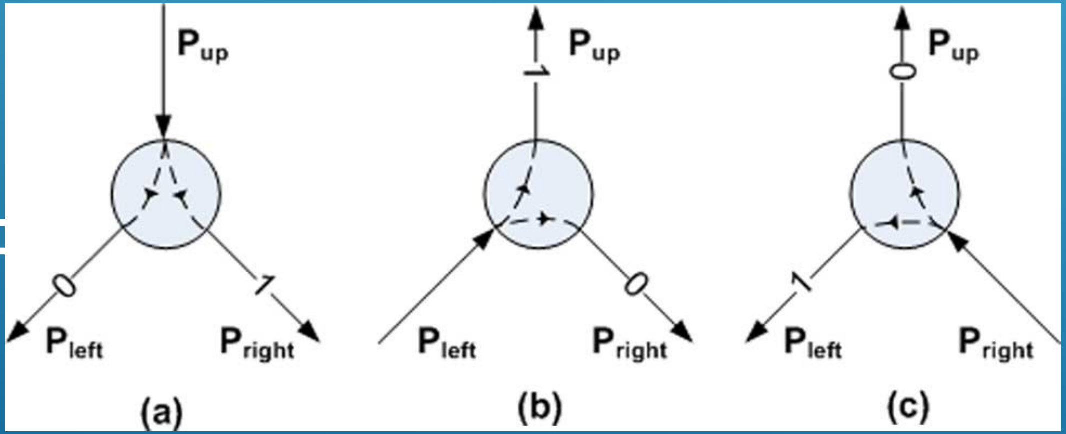

Source
A
C
N_j
E
I
J
N_q
10
F
K
B
N
101100110
1011001101001
G
10110
Destination
101100110100101
D
N_i
L
O
H
M
N_r

- Creating Virtual Topology



- Binary Routing

CONCEPT OF



ADDRESSING AND ROUTING IN OE BASED TREES

Discard common MSBs

0	0	0	1	Source S-ARTAG

0	1	1	1	0	Destination S-ARTAG

Complement Source S-ARTAG except MSB

0	1	0

Discard MSB of Destination S-ARTAG

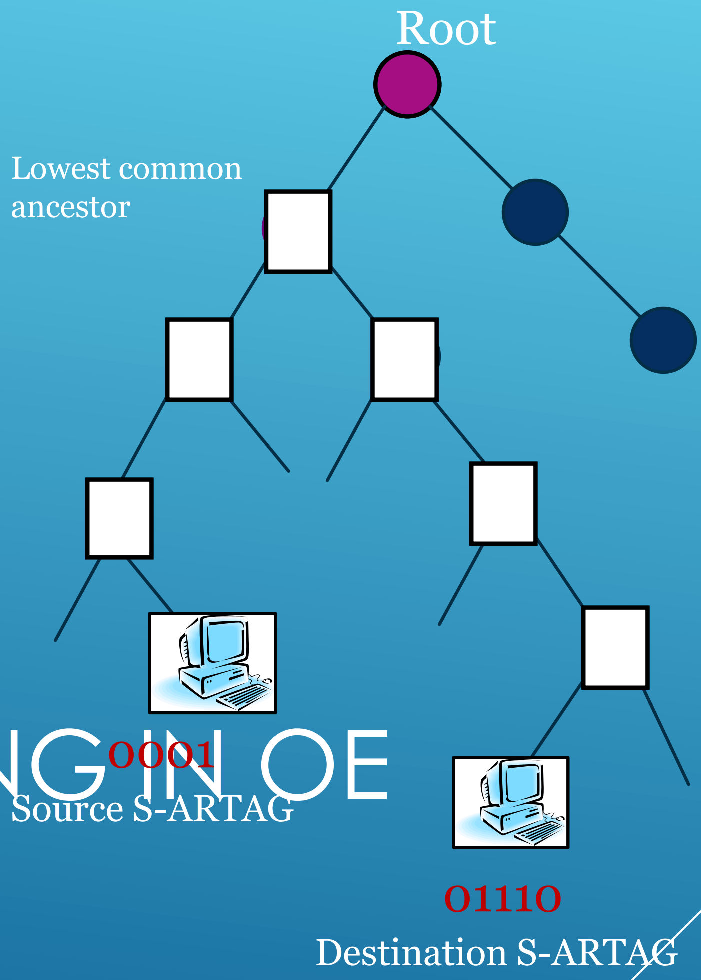
1	1	0

Reverse Destination S-ARTAG

0	1	1

Final R-ARTAG

0	1	1	0	1	0



Omnipresent Ethernet another avatar of MPLS and SDNs

Ethern
frame
with c
et
r

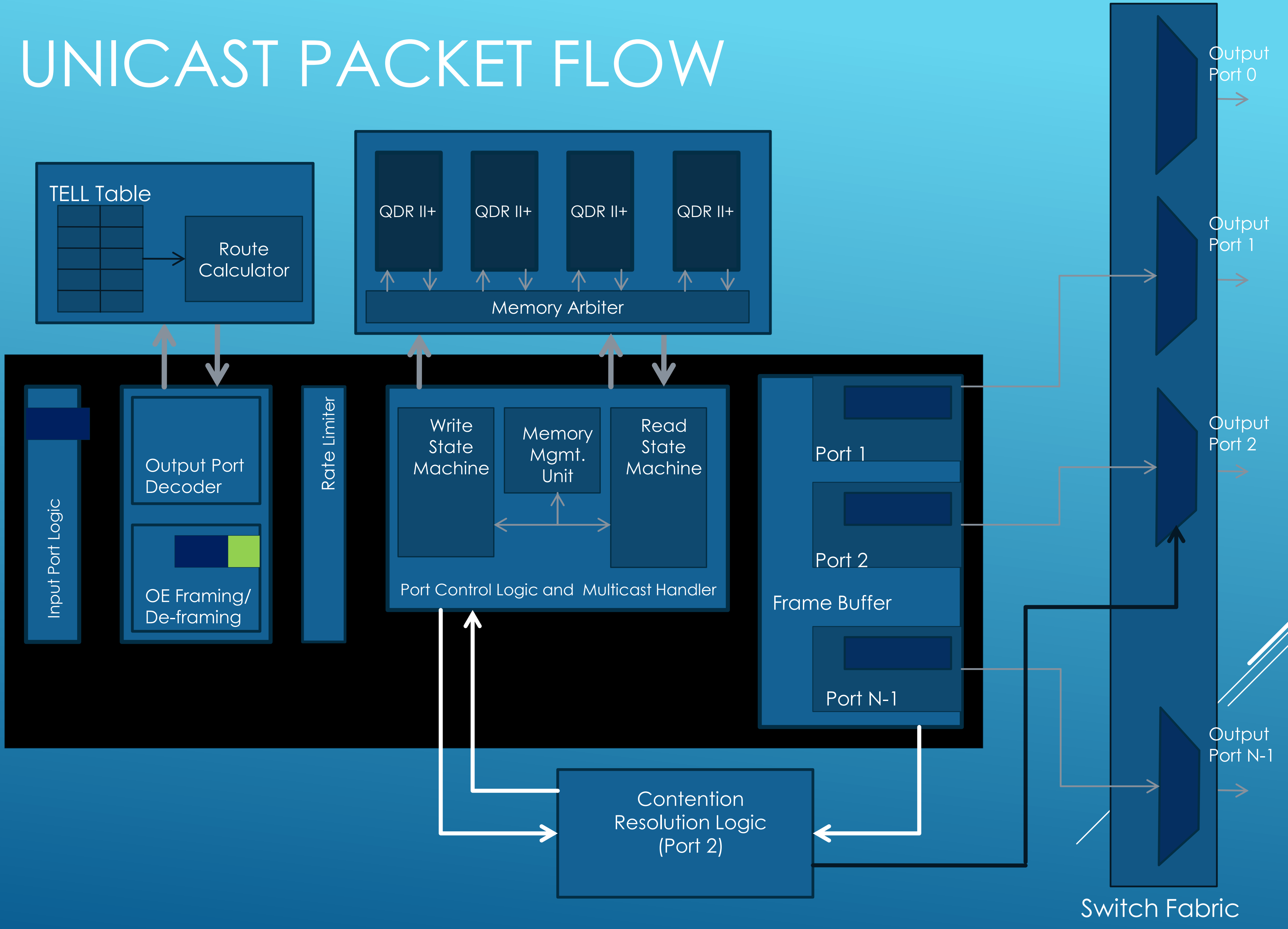
dat
a

Identifier	Binary Tag
IP address IPV4	010001001
IP address IPV6	100101010
MAC	1010
Port	111010
S/CTAG	0010101010 1

Ethernet data mapped onto OE packet – Ethernet header used for L2 protocol.


ECR - 100
MGMT
PWR
Electronics Corporation of India Ltd
ELECTRONICS CORPORATION OF INDIA LTD.
ECR-1000
ECR-1010

UNICAST PACKET FLOW



THE CESR HARDWARE

Fig 1. The Carrier Ethernet Switch Router(CESR) hardware

THE V SERIES CORE ROUTER, LONG HAUL TRANSPORT, TUNABLE WDM SUPPORT WITH CARRIER ETHERNET

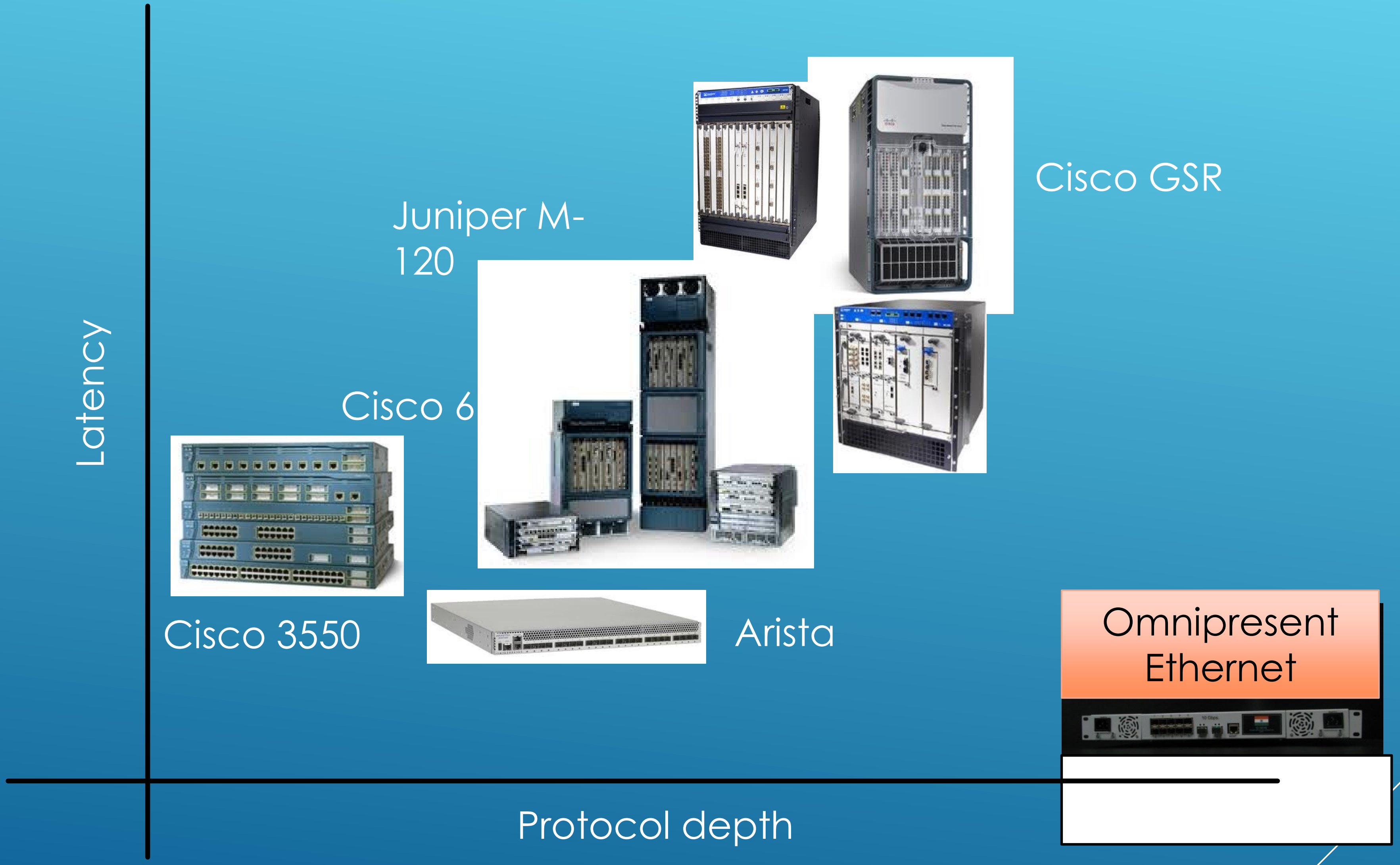
- State-of-the-art transport solution.
- 1000 km reach without regen.
- 96 Gbps cross connect
- OTN as ODU2 compliant.
- 1000 FEC entries
- PseudoWire emulation.
- Deep buffers for packet processing
- 3-5 microsecond latency.
- Multicast, 4 level QoS support.
- Dense Wavelength Division Multiplexing technology for super fiber utilization.
- Applications: metro transport, regional transport, multi-Gigabit router, National Knowledge Network transport, enterprise backbone and Carrier Ethernet.

STATISTICS

	O1000 – LX240T (IPv4)	O1000 – LX365T (IPv6)	O1010 – LX365T	O100 – LX[illegible]
FPGA device Utilization	92%	67%	71%	71%
BRAM Utilization	59%	61%	74%	74%
Lines of code (VHDL)	1,28,405*	1,28,405*	68,302**	68,302**
Lines of code (NMS)	16,142	16,142	16,142	16,142
Lines of code Web based NMS	50,000 +	50,000 +	50,000 +	50,000 +
PCB stats	R ~ 500, C ~ 1000	R ~ 500, C ~ 1000	R ~ 300, C ~ 700	R ~ 700, C ~ 2000
	135 different components	135 different components	107 different components	176 different components
Total components	1500+	1500+	2200+	850

*O1000 code statistics includes test code
**O100 and O1010 code statistics exclude common files from O1000.

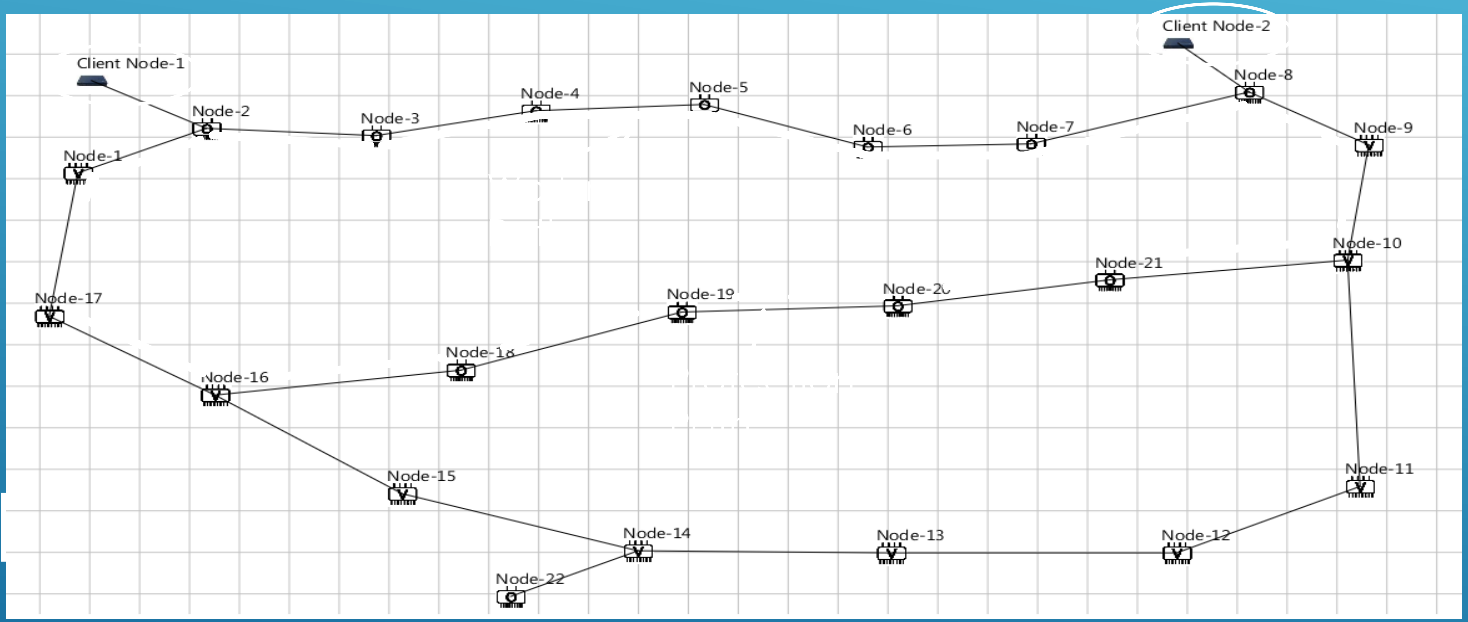
COMPARISON


Latency
Juniper M-120
Cisco GSR
Cisco 6
Cisco 3550
Arista
Omnipresent Ethernet
Protocol depth

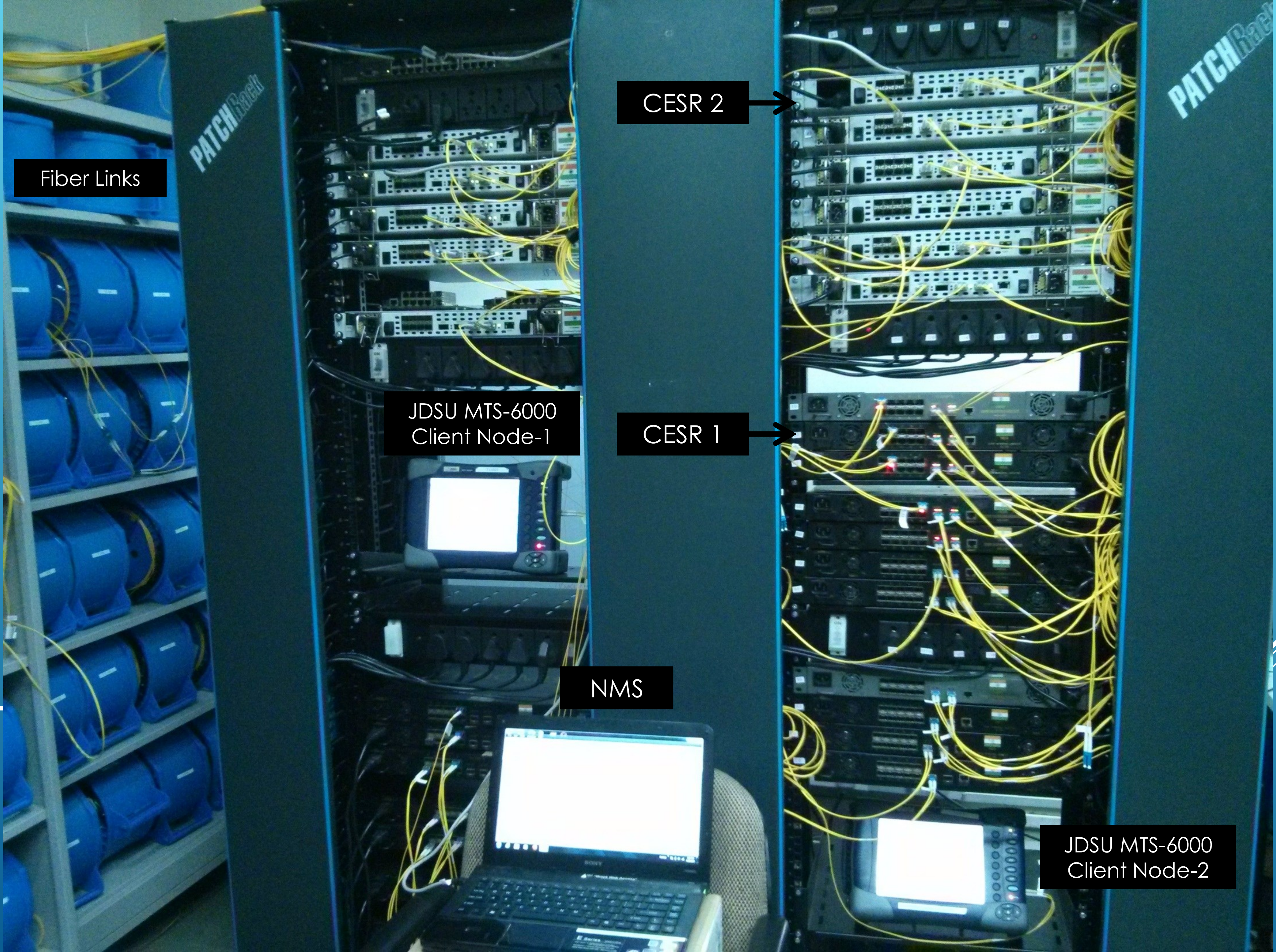
CESRS DEVELOPED

- F
- C
- DWDM optics supported
- E-LINE, E-LAN service support for MAC,IPv4,IPv6, CTAG, STAG, port based identifiers

- RailTel Western Region Network
  - 22 nodes
  - 23 links



EXPERIM


Fiber Links
CESR 2
JDSU MTS-6000
Client Node-1
CESR 1
NMS
JDSU MTS-6000
Client Node-2

TES